Economic and Healthcare Related Determinants of Infant Health at Birth

Samuel I. Watson

Warwick Medical School

University of Warwick

This dissertation is submitted for the degree of

Doctor of Philosophy

February 2015

Table of contents

Table of contents i

List of figures vii

List of tables ix

Glossary xxi

1 Introduction 1

2 Data Sources and Key Variables 19

2.1 Introduction 19

2.2 The National Neonatal Research Database 20

2.2.1 Missing Data 21

2.2.2 Neonatal Unit Characteristics 23

2.2.3 Clinical Variables 27

2.2.4 Maternal Residence 34

2.2.5 Maternal Variables 35

2.3 Population data 35

2.3.1 Geographical Areas 36

2.3.2 Socio-economic Deprivation 36

2.3.3 Unemployment Rate 39

2.3.4 Area Level Birth Rate and Birth Weight 39

2.4 National Health Service data 40

2.4.1 Programme Budgeting Data 42

2.4.2 Reference Costs 43
2.5 Unit Profile Survey 2011 46

3 Literature Review 47
3.1 Methods 48
3.2 Results from the Literature Review 49
3.2.1 Review of Findings of Included Studies 54
3.2.2 Research Methods 63
3.3 Discussion 68

4 The effect of volume and designation of care 71
4.1 Introduction 71
4.2 Sample selection 74
4.3 Variables 75
4.3.1 Volume and Designation of Care 75
4.3.2 Outcomes 76
4.3.3 Covariates 77
4.3.4 Instruments 77
4.4 Econometric specification 78
4.5 Results 81
4.5.1 Summary statistics 81
4.5.2 Standard Logistic Regression 83
4.5.3 Instrumental Variable Validity 86
4.5.4 Instrumental Variable Logistic Regression 88
4.6 Robustness 91
4.7 Summary and Conclusions 92

5 Neonatal health care expenditure 100
5.1 Background 103
5.1.1 The National Health Service in England 103
5.1.2 Previous Literature 105

5.1.3 A Comment on Claxton et al. (2013) . . . 106
5.1.4 Other Literature . . . 110
5.2 Sample Selection . . . 111
5.3 Model and variables . . . 111
5.3.1 Dependent Variable . . . 114
5.3.2 Control variables . . . 114
5.3.3 Estimation . . . 114
5.3.4 Other Issues . . . 115
5.4 Results . . . 117
5.4.1 Summary Statistics . . . 117
5.4.2 Instrument Validity . . . 118
5.4.3 Main Results . . . 119
5.5 Robustness and Sensitivity . . . 121
5.6 Exploring the Expenditure Effect . . . 123
5.7 Cost-Effectiveness Threshold . . . 128
5.7.1 Cost Per Life Gained . . . 128
5.7.2 Cost per life year gained . . . 128
5.7.3 Additional Costs at the Margin . . . 129
5.7.4 Heterogeneous Effects by Unit Volume . . . 132
5.8 Discussion and Conclusions . . . 133

6 One to one nursing 137

6.1 Background . . . 139
6.1.1 Clinical Guidelines . . . 139
6.1.2 Previous literature on Nurse to Patient Ratios in Neonatal Healthcare . . . 140
6.2 Conceptual Model . . . 142
6.3 Variables . . . 145
6.3.1 One to one nursing . . . 145
6.3.2 Outcomes . . . 149

6.3.3 Control variables 149
6.4 Econometric Model and Sample 149
6.4.1 Analysis sample 154
6.4.2 Definition of one to one nursing 155
6.5 Results 156
6.5.1 Summary Statistics 156
6.5.2 IV first stage 157
6.5.3 Main results 160
6.5.4 Robustness and Sensitivity 160
6.6 Discussion and Conclusions 162

7 Economic Conditions and Infant Health 167
7.1 Theoretical Background 171
7.2 Data and Variables 173
7.2.1 Outcome variables 174
7.3 Specification and Estimation 175
7.3.1 Data issues 176
7.3.2 Model 1: Pooled Cross Section 177
7.3.3 Model 2: Panel Fixed Effects 178
7.3.4 Model 3: Selection Effects 178
7.3.5 Model 4: Measurement Error 181
7.3.6 Other Issues 183
7.4 Results 183
7.4.1 Descriptive statistics 183
7.4.2 Admissions 186
7.4.3 Birth-weight 187
7.4.4 Clinical outcomes 188
7.4.5 Birth rate 191
7.5 Robustness and Sensitivity 191
7.5.1 Sample selection 192

7.5.2 Instrumental variables results 193
7.6 Extensions 194
7.6.1 Results by socio-economic status 194
7.6.2 Economic status 197
7.6.3 Gender 200
7.7 Discussion and conclusions 201

8 Normative Issues 204
8.1 The Social Discount Rate 206
8.1.1 Time Preference 207
8.1.2 Marginal Utility of Consumption 208
8.2 Measuring benefits 210
8.3 Distributional Justice 212
8.4 Centralisation of Neonatal Healthcare Services 218

9 Conclusions 222

References 233

Appendix A Addition Information for Chapter 3: Literature Review 249
A.1 Table of Results 249

Appendix B Addition Information for Chapter 4: Neonatal Unit Volume and Designation 266
B.1 Regression Results 266

Appendix C Addition Information for Chapter 5: Neonatal Healthcare Expenditure 271
C.1 A Simple Example of the Issue With Aggregate Expenditure Data . . 271
C.2 Regression Results 273

Appendix D Addition Information for Chapter 7: One to One Nursing 276
D.1 Regression Results 276

Appendix E Addition Information for Chapter 7: Economic Conditions and Infant Health 278

E.1 Further Details on the Econometric Specification 278

E.1.1 Model 3: Correction for Sample Selection Bias 278

List of figures

1.1 UK-WHO Neonatal and Infant Close Monitoring preterm growth chart 3

1.2 Model of pathways of social influences on health 16

2.1 Map of neonatal units in England by neonatal network 26

2.2 The distribution of annual care days provided by unit designation . . . 28

2.3 Gestational age 30

2.4 Average length of stay by birth weight 35

2.5 The structure of the NHS in England, 2003-13. 41

2.6 Distribution of estimated unit costs of neonatal healthcare by care level 45

3.1 Crude mortality rates by unit level and volume of VLBW admissions for units in California 1997-2002, from Chung et al., (2010) 53

3.2 Estimated risk of mortality by maximum occupancy at admission with 95% confidence interval from Tucker (2002) 55

3.3 The estimated odds ratio for hospitals with large and small NICUs and large and small delivery units adjusted for a number of baby covariates, a large NICU was defined as having $\geq$36 VLBW admissions annually and a large delivery unit was defined as having $\geq$1,000 births a year, from Bartels et al. (2006) 57

5.1 PCT distances from target allocations 105

5.2 Correaltion between average care day costs and factor inputs. 125

6.1 Correlation between the ratio of whole time equivalent nurses to intensive care care days and the ratio of one to one nursing days and intensive care days. 148
6.2 Relationship between infant characteristics at birth and receipt of one to one nursing 150
6.3 Variation in supply of one to one nursing and intensive care days . . . 154
6.4 Variation in one to one nursing over time 158

7.1 Total numbers of lagged JSA claimants, lagged income support claimants, and births between 2006-11 in England 185
7.2 Birth weight quantile-quantile plots 188
7.3 Comparison of the proportion of individuals in 'professional' occupations and the proportion of individuals in 'working class' occupations for each MSOA 197

List of tables

2.1 Descriptive statistics of the NNRD data 22

2.2 Percentages of missing data by variable and year for selected variables in the NNRD 24

2.3 Definitions of gestational age 31

2.4 Super Output Areas in England 36

2.5 Descriptive statistics of infants by quintile of deprivation 38

2.6 Summary statistics for numbers of births by birth weight and proprtion of admissions 40

2.7 Fertility and birth statistics for live births in England and Wales 40

2.8 Summary statistics for programme budegting data 44

3.1 Results from literature review searches 50

3.2 The number of studies reporting various unit characteristics and outcomes 51

4.1 Summary statistics of the sample 82

4.2 Odds ratios from standard logistic regression 84

4.3 Average partial effects from logistic regression 85

4.4 Summary statistics by levels of the instrumental variables 87

4.5 Odds ratios from instrumental variable logisitic regression 89

4.6 Average partial effects from instrumental variable logisitic regression 90

4.7 Results from sensitivity analysis. Infants with missing data excluded from analyses. 96

4.8 Results from sensitivity analysis. Morbidity outcomes. 97

4.9 Results from sensitivity analysis. Alternative measures of volume. . . 98
4.10 Results from sensitivity analysis. Alternative measures of volume. . . 99

5.1 Summary statistics of the sample 117
5.2 Regression results treating expenditure as exogenous 120
5.3 Regression results treating expenditure as endogenous 120
5.4 Regression results treating expenditure as endogenous without inverse probability weighting estimation. 121
5.5 Regression results treating expenditure as endogneous excluding level one units 122
5.6 Regression results treating expenditure as endogenous excluding data from 2012/13 122
5.7 Regression results treating expenditure as endogenous and estimated separately for high and low volume neonatal units 127
5.8 Estimated cost per statistical life for infants born in 2011/12 128
5.9 Estimated cost per statistical life for infants born in 2011/12 133

6.1 Baby level summary statistics comparing babies to have been provided one to one nursing to those that were not 156
6.2 Individual level summary statistics 157
6.3 Results from the first stage regression 159
6.4 Estimated effect of one to one nursing rate on mortality rate 161
6.5 Estimated effect of one to one nursing rate on mortality rate 161
6.6 Estimates from sensitivity analyses 163

7.1 Summary statistics for included/observed and excluded/unobserved MSOAs. 179
7.2 Summary statistics for MSOAs by year 184
7.3 Effect of the local unemployment rate on admissions to neonatal care . 186
7.4 Effect of unemployment on infant birth weight 187
7.5 Effect of local unemployment rate on the proportion of births admitted to neonatal healthcare with various conditions 190

7.6 Effect of unemployment on the birth rate 191
7.7 Effect of local unemployment on admissions to neonatal care, various methods for selecting observed MSOAs. 192
7.8 Effect of local unemployment on admissions to neonatal care with lagged local unemployment rate as instrument for local unemployment rate. 193
7.9 Effect of local unemployment on infant health disaggrated by quintile of the Index of Multiple Deprivation and social class 198
7.10 Summary statistics for areas with high and low average unemployment in 2006-11 using data from census 2011. 199
7.11 Effect of local unemployment on admissions to neonatal care, different measures of unemployment rate 200

8.1 Scenario 1: Life Years Resulting from Two Alternative Policies 216
8.2 Scenario 2: Effects of a Centralisation Policy 219

A.1 Results from the literature review 250

B.1 Parameter estimates from standard logistic regression for admission to high volume neonatal unit at the place of birth for very preterm infants 267
B.2 Parameter estimates from standard logistic regression for admission to level three neonatal unit at the place of birth for very preterm infants . 268
B.3 Parameter estimates from instrumental variable logistic regression for admission to high volume neonatal unit at the place of birth for very preterm infants 269
B.4 Parameter estimates from instrumental variable logistic regression for admission to level three neonatal unit at the place of birth for very preterm infants 270

C.1 Expenditure on medicine and the number of coughs in three different scenarios 272
C.2 Regression results treating expenditure as exogenous 274

C.3 Regression results treating expenditure as endogenous 275

D.1 Estimated effect of one to one nursing rate on mortality rate 277

I would like to dedicate this thesis to my parents.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author except in the case outlined below:

Chapter 4 was conducted in collaboration with Wiji Arulampalam (WA), Stavros Petrou (SP), Neil Marlow (NMa), Elizabeth Draper (ED), Neena Modi (NMo), Shalini Santhakumaran (SS), and Andrei Morgan (AM). SW conceived the study; SW, WA, and SP contributed to developing the econometric methodology for the study; SW prepared the data for analysis; SW, WA, SP, NMa, AM, ED, and NMo contributed to covariate selection and interpretation of the results; SS managed the extraction and cleaning of NNRD variables; SW prepared the first draft of the chapter; this and all subsequent drafts were reviewed and revised by all authors; all authors approved the final version submitted.

Parts of this thesis have been published by the author: Watson, S. I., Arulampalam, W., Petrou, S., Marlow, N., Morgan, a. S., Draper, E. S., and Modi, N. (2014). The effects of designation and volume of neonatal care on mortality and morbidity outcomes of very preterm infants in England: retrospective population-based cohort study. BMJ Open, 4(7), e004856–e004856. doi:10.1136/bmjopen-2014-004856

Samuel I. Watson

February 2015

Acknowledgements

I would like to thank the many individuals who contributed both academically and personally and who made this thesis possible. Firstly, I would like to thank my supervisors, Stavros Petrou and Wiji Arulampalam, for their continued insight and patience without whom I could not have completed this work. The work presented for this thesis was conducted as part of the Neonatal Economic Staffing and Outcomes Project (NESCOP). The NESCOP group comprises Wiji Arulampalam (University of Warwick), Liz Draper (University of Leicester), Neil Marlow (University College London), Neena Modi (Imperial College, London), Andrei Morgan (Imperial College, London), Stavros Petrou (University of Warwick), Shalini Santhakumaran (Imperial College, London) and Sam Watson (University of Warwick), with support from Bliss, in particular Andy Cole. I thank the whole NESCOP group for valuable input to all the chapters of this thesis. I spent much of the duration of the past three years working with the Neonatal Data Analysis Unit who have frequently aided me with the analyses presented in this thesis. In particular, I thank Nicola Fitz-Simon and Shalini Santhakumaran for invaluable advice on statistical issues, Eugene Statnikov and Daniel Gray for assistance with the data, Richard Colquhoun, Neonatal Data Analysis Unit manager, and Surbhi Shah for administrative support. I also thank Kate Bennett, Helen Kirrane, and the Steering Board of the Neonatal Data Analysis Unit.

I would like to thank the Department of Economics, University of Warwick and Bliss for the funding that made this work possible.

I would not have been able to make it to the end of this thesis without invaluable support from my friends and family. I am most grateful to all my friends for their support, but I want to extend particular thanks to Tom White who, over the last three

years, put up with more of my frustrations than anyone else, and James Cowley, Luke Pagarani, Phil Smith, Cameron Westwood, Clare Stimpson, Adam Hamblin, Pablo Geijo de Celis, Sanjay Patel, and Richard McLaverty who have all been a source of support and priceless distraction.

All neonatal units whose data were included in this study were contacted as part of the NESCOP survey and offered an 'opt out' of service evaluation that made use of the NNRD data. I would like to thank the contributing neonatal units and clinical leads:

Matthew Babirecki, Airedale General Hospital; Andrew Gallagher, Alexandra Hospital; D Manning, Arrowe Park Hospital; Tim Wickham, Barnet Hospital; MA Saeed, Barnsley District General Hospital; Nawfal Sharief, Basildon Hospital; Ruth Wigfield, Basingstoke & North Hampshire Hospital; Nigel Brooke, Bassetlaw District General Hospital; Raghavan Kadalraja, Bedford Hospital; Julie Nycyk, Birmingham City Hospital; Mike Watkinson, Birmingham Heartlands Hospital; Imogen Morgan, Birmingham Women's Hospital; Sam Oddie, Bradford Royal Infirmary; R.N Mahesh Babu, Broomfield Hospital; Eleri Crosbie, Calderdale Royal Hospital; Tim Wickham, Chase Farm Hospital; Gary Hartnoll, Chelsea & Westminster Hospital; Aiwyne Foo, Chesterfield Royal Hospital; Sarah Dalton,Colchester General Hospital; Graham Whincup, Conquest Hospital; Stephen Brearey, Countess of Chester Hospital; John Chang, Croydon University Hospital (Mayday); Mahfud Ben-Hamida, Cumberland Infirmary; Abdul Hasib, Darent Valley Hospital; S Subramania Krish Namoorthy, Darlington Memorial Hospital; Julian Larson, Derriford Hospital; Kallinath Shyamannr, Dewsbury & District Hospital; Pauline Adiotomre, Diana Princess of Wales Hospital; Nigel Brooke, Doncaster Royal Infirmary; Phil Wylie, Dorset County Hospital; Vivien Chan, Ealing Hospital; K Aboul Khader, East Surrey Hospital; Graham Whincup, Eastbourne District General Hospital; Ruth Shephard, Epsom General Hospital; Sri Nagesh Panasa, Fairfield general; JP Wyllie, Friarage Hospital; Abdus Mallik, Frimley Park Hospital; Owen Galt, Furness General Hospital; R C de Boer, George Eliot Hospital; Jennifer Holman, Gloucestershire Royal Hospital; Mike Watkinson, Good Hope Hospital; Stanley Zengeya, Great Western Hospital; Morag Campbell, Guy's & St Thomas' Hospital; D Gillies, Harrogate District Hospital; Andrew Butter-

fill, Hereford County Hospital; Michele Cruwys, Hillingdon Hospital; Hilary Dixon, Hinchingbrooke Hospital; Kate Costeloe, Homerton Hospital; Eleri Adams, Horton Hospital; Chris Wood, Hull Royal Infirmary; Matthew James, Ipswich Hospital; JP Wyllie, James Cook University Hospital; Vasantha Jayalal, James Paget Hospital; Eleri Adams, John Radcliffe Hospital; Pratibha Rao, Kettering General Hospital; B K Sharma, King George Hospital; Abhi Gkeenga, King's College Hospital; Vibert Noble, King's Mill Hospital; Dwight Lindo, Kingston Hospital; Meera Lama, Lancashire Women and Newborn Centre; Lawrence Miall, Leeds Neonatal Service; Andrew Currie, Leicester General Hospital; Andrew Currie, Leicester Royal Infirmary; A Thirumurugan,L eighton Hospital; Sudhakar Rao, Lincoln County Hospital; Andy Raffles, Lister Hospital; CW Yoxall, Liverpool Women's Hospital; P Sivakumar, Luton & Dunstable Hospital; Ignatios Losa, Macclesfield District General Hospital; Hamudi Kisat, Maidstone; B Jan Muhamad, Manor Hospital; Aung Soe, Medway Maritime Hospital; Jasper Katumba, Milton Keynes Foundation Trust Hospital; Babu Kumararatne, New Cross Hospital; Vimala Gopinathan, Newham General Hospital; Mark Dyke, Norfolk & Norwich University Hospital; Yonas Cherinet ,North Devon District Hospital; Sri Nagesh Panasa, North Manchester General Hospital; Lesley Alsford, North Middlesex University Hospital; Fiona Thompson, Northampton General Hospital; Paul Mannx, Northwick Park Hospital; Stephen Wardle, Nottingham City Hospital; Stephen Wardle, Nottingham University Hospital; Tim McBride, Ormskirk District General Hospital; Seif Babiker, Peterborough City Hospital; Sudhakar Rao, Pilgrim Hospital; Minesh Khashu, Poole Hospital NHS Foundation Trust; Than Soe, Princess Alexandra Hospital; Mike Hall, Princess Anne Hospital; Philip Amess, Princess Royal Hospital; Shaun Walter, Princess Royal University Hospital; Andy Raffles, Queen Elizabeth II Hospital; CR Groves, Queen Alexandra Hospital; Lidia Tyszczuk, Queen Charlotte's Hospital; Susan Rubin, Queen Elizabeth Hospital; Shaun Walter, Queen Elizabeth Hospital; Dennis Bosman, Queen Elizabeth Hospital (Gateshead); David Long, Queen Elizabeth the Queen Mother Hospital; Azhar Manzoor, Queen's Hospital; B K Sharma, Queen's Hospital; Amanda Ogilvy-Stuart, Rosie Maternity Hospital; Christine Harrison, Rotherham District General Hospital; Vibha Sharma, Royal Albert Edward Infir-

mary; GJ Boson, Royal Berkshire Hospital; M Yadan, Royal Bolton Hospital; Paul Munyard, Royal Cornwall Hospital; John McIntyre, Royal Derby Hospital; Vaughan Lewis, Royal Devon & Exeter Hospital; D Schapira, Royal Hampshire County Hospital; Owen Galt, Royal Lancaster Infirmary; Natasha Maddock, Royal Oldham Hospital; Richa Gupta, Royal Preston Hospital; Alison Moore, Royal Shrewsbury Hospital; Charles Godden, Royal Surrey County Hospital; Philip Amess, Royal Sussex County Hospital; Steve Jones, Royal United Hospital; Alan Fenton, Royal Victoria Infirmary; Anande Mohite, Russells Hall Hospital; Jonathan Moise, Salford Royal; Shirley Kinsey, Salisbury District Hospital; Mazen Qunibi, Scarborough General Hospital; Pauline Adiotomre, Scunthorpe General Hospital; Rob Bolton, South Tyneside District Hospital; Arfa Khan, Southend Hospital; Richard Wach, Southmead Hospital; Sandra Calvert, St George's Hospital; Ruth Shephard, St Helier Hospital; Anthony Emmerson, St Mary's Hospital; Peter Chow, St Mary's Hospital; Sian Butterworth, St Mary's Hospital; James Tooley, St Michael's Hospital; Peter Reynolds, St Peter's Hospital; Avril Garg, St Richard's Hospital; KK Tewary, Staffordshire General Hospital; Carrie Heal, Stepping Hill Hospital; Gail Whitehead, Stoke Mandeville Hospital; Majd Abu-Hard, Sunderland Royal Hospital; Jacqueline Birch, Tameside General Hospital; Rebecca Mann, Taunton & Somerset Hospital; Elizabeth Pilling, The Jessop Wing; Vivienne van Someren, The Royal Free Hospital; Rainer Ebel, The Royal London Hospital; Stefan Schulee, Torbay Hospital; Hamudi Kisat, Tunbridge Wells Hospital; Mark Sellwood, University College Hospital; Kate Blake, University Hospital Coventry; Jauro Kuna, University Hospital Lewisham; S Subramania Krish Namoorthy, University Hospital of North Durham; Kate Palmer, University Hospital of North Staffordshire; Ian Verber, University Hospital of North Tees; Faisal Al-Zidgali, University Hospital of South Manchester; AJ Rowlands, Victoria Hospital; Joan Oliver, Wansbeck General Hospital; Delyth Webb, Warrington Hospital; R B Acharya, Warwick Hospital; Chaniyil A Ramesh,Watford General Hospital; Mahfud Ben-Hamida, West Cumberland Hospital; Hashir Ariff, West Middlesex University Hospital; Ian Evans, West Suffolk Hospital; Rekha Sanghavi, Wexham Park Hospital; Caroline Sullivan, Whipps Cross University Hospital; Laweh Amegavie, Whiston Hospital; ER Broad-

hurd, Whittington Hospital; David Long, William Harvey Hospital; Andrew Gallagher, Worcestershire Royal Hospital; Anil Garg, Worthing Hospital; Megan Eaton, Yeovil District Hospital; Guy Millman, York District Hospital.

Abstract

This thesis analyses the effects of various structural and organisational characteristics of specialist neonatal units on the clinical and economic outcomes of infants treated within them. Data are utilised from the National Neonatal Research Database (NNRD) which is extracted from the electronic patient records of all infants admitted to the vast majority of neonatal units in England over the period 2006-13 along with national healthcare expenditure and demographic data. Firstly, I examine the effects of neonatal unit volume and designation on infant clinical outcomes. In 2003, neonatal units in England and Wales were re-organised into networks to facilitate access to high level and volume neonatal units for the sickest infants as infants treated in these units had previous been shown to be at less risk of adverse outcomes. No previous studies have examined the effects of neonatal unit volume and designation in such a networked setting. Secondly, I estimate the effect of neonatal healthcare expenditure on the risk of mortality, and in so doing determine the cost-effectiveness of neonatal healthcare. Thirdly, I analyse the effect of nurse to patient ratios in neonatal intensive care on the risk of mortality, recent evidence has demonstrated that neonatal units are often understaffed with respect to clinical guidelines, yet little is known about the consequences of this on infant clinical outcomes. Finally, I explore the effect of local economic conditions at the time of conception on infant health at birth. The number of admissions to neonatal specialist healthcare units has increased in recent years to approximately 10% of all live births. Understanding the mechanisms underlying this increase is important both for healthcare capacity planning and also development of policies aimed at improving infant health at birth. The results in this thesis support policies aimed at increasing the proportion of infants born in hospitals with high volume neonatal units along with an increased provision of resources for neonatal healthcare.

Glossary

2SLS Two Stage Least Squares.

ANNP Advanced Neonatal Nurse Practitioner.

BAPM British Association of Perinatal Medicine.

BPD Bronchopulmonary Dysplasia.

CBA Cost Benefit Analysis.

CEA Cost Effectiveness Analysis.

CUA Cost Utility Analysis.

Early Neonatal Mortality Death within the first seven days of post-birth.

ELBW Extremely Low Birth Weight: born at $\leq$1,000g.

FE Fixed Effects.

Gestational Age The length of a pregnancy, commencing at the last menstrual period, typically measured using obstetric ultrasonography.

GMM Generalised Method of Moments.

HDC High Dependency Care.

IC Intensive Care.

ICER Incremental Cost Effectiveness Ratio.

In Hospital Mortality Death between admission and discharge from a neonatal unit.

LBW Low Birth Weight: born at $\leq$2,500g.

MCN Managed Clinical Network.

NEC Necrotising Enterocolitis.

Neonatal Mortality Death within the first 28 days of post-birth.

NICE National Institute of Health and Care Excellence.

NICU Neonatal Intensive Care Unit.

NNRD National Neonatal Research Database.

OLS Ordinary Least Squares.

PCT Primary Care Trust.

ROP Retinopathy of Prematurity.

SC Special Care.

Small for Gestational Age Birth weight is below the 10th percentile for gestational age.

VLBW Very Low Birth Weight: born at $\leq$1,500g.

Chapter 1

Introduction

Over the last forty years there has been a dramatic reduction in the infant mortality rate in developed nations. Between 1980 and 2012 the mortality rate in children aged under one year declined in England and Wales from 12.0 deaths per 1,000 live births to 4.0 deaths per 1,000 live births (Office for National Statistics, 2014). A major contributory factor has been the substantial advances in neonatal medicine, which have led to a reduction in the neonatal mortality rate (death within 28 days post birth) from 7.7 to 2.8 deaths per 1,000 live births over the same period (Office for National Statistics, 2014). Despite the improvement in neonatal care, the field still faces a number of challenges. The number of preterm births has increased in recent years, for example, there were 11% more preterm singleton births in 2012 than in 1985 (Norman et al., 2009; Office of National Statistics, 2012). Moreover, as survival rates increase for very preterm babies, the absolute numbers of individuals with diseases associated with prematurity would be expected to increase (Iams et al., 2008). As such the focus of neonatal healthcare policy in recent years has been to ensure that vulnerable infants are able to access and receive adequate, appropriate care. This has presented various challenges to policy makers regarding the optimal organisation and resourcing of neonatal specialist care. The aim of this thesis is to provide empirical evidence centred on a rich and novel data source covering neonatal admissions in England to aid policy makers with the organisation, planning, and resourcing of neonatal healthcare.

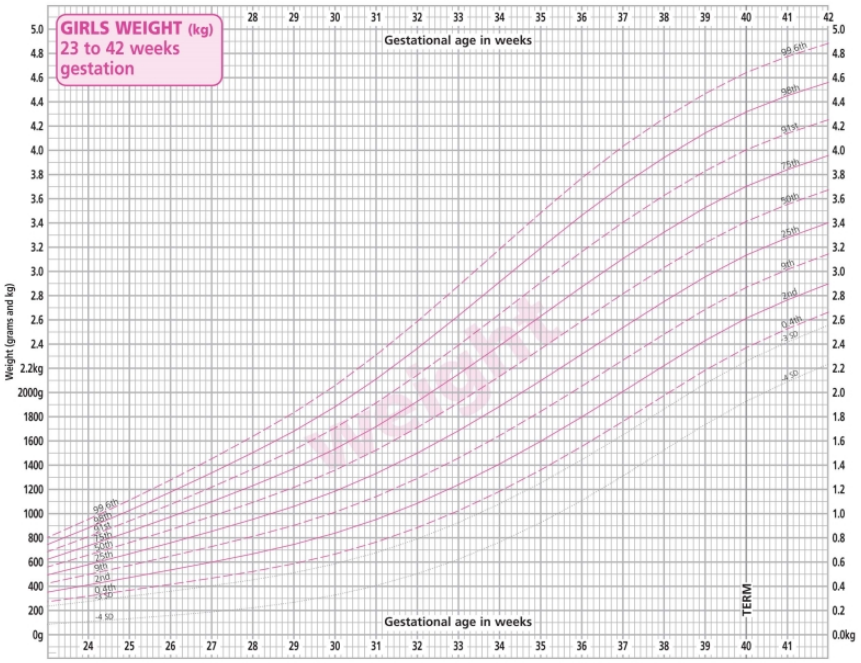
Neonatal medicine is a relatively recent medical speciality. While the specialised

care of newborn infants has been practised for over one hundred years, modern neonatology, as it is currently recognised, has only existed for the last fifty years. The term 'neonatology' was coined in 1960 by Alexander Schaffer, and it was around this time that clinical definitions of preterm birth (e.g. less than 37 weeks gestation—see Section 2.2.3.1 of Chapter 2 for the current definitions) were developed (Philip, 2005). In the ensuing decades a number of technical advances were made that led to dramatic increases in the perinatal survival rate, particularly among very low birth weight (VLBW; <1,500g birth weight) or very preterm (born at less than 33 weeks gestation) infants, for example, the survival rate of infants born at less than 1,000g was 5% in the 1960s and 73% in the 2000s (Behrman, 1971; Latini et al., 2013).[1] These advances include, but are by no means limited to: thermoregulation (ensuring infants maintain the correct body temperature); improved nutrition, both in terms of its composition and administration; establishment of growth norms (Figure 1.1 shows a preterm growth chart); respiratory support and ventilation; and cardiopulmonary support (Philip, 2005). The technological advances have also necessitated complementary improvements in the knowledge and skills of the labour force. In the United States, the American Academy of Pediatrics has published the "Guidelines for Perinatal Care" since 1983, which is now in its seventh edition (American Academy of Pediatrics Committee on Fetus and Newborn and American College of Obstetricians and Gynecologists Committee on Obstetric Practice, 2012). In the United Kingdom, neonatal care has been a recognised nursing sub-speciality since 1992 (and from the 1970s in the US) (Hall et al., 1992).

Modern neonatal units provide care for a wide variety of different conditions of the newborn for which staff members must be equipped. These range from congenital anomalies, which are structural deformities of the fetus, to infants born extremely prematurely. Care on a neonatal unit is provided by a wide variety of different staff members. Neonatologists are the clinicians responsible for the provision of care on a neonatal unit; neonatology is a sub-speciality of pediatrics. Neonatologists are supported by a number of other staff members including neonatal nurses, advanced neona-

[1] A birth weight of 1,000g or less is classified as extremely low birth weight. The average birth weight of live births in the United Kingdom is currently 3,300g with only 0.5% of all live births being born at 1,000g or less (Office for National Statistics, 2014).

Fig. 1.1 UK-WHO Neonatal and Infant Close Monitoring preterm growth chart



WHO = World Health Organisation. Source: Royal College of Pediatrics and Child Health (2013)

tal nurse practitioners (ANNPs), and health care assistants (HCA). Neonatal nurses are nurses that have obtained additional training in neonatal care and who provide day to day care for the infants admitted to neonatal units, including changing nappies, administering blood products and intravenous fluids, and monitoring blood oxygen levels. ANNPs are nurse consultants who prepare and manage treatment plans for neonates and provide a managerial input to the unit. HCAs are support workers who provide information to parents, aid with aspects of neonatal care such as breastfeeding initiation, as well as conduct general administrative duties. There are a number of other members of staff who provide an important input to neonatal health care including family support workers, physiotherapists, and transport teams.

There are a number of organisational characteristics that healthcare policy makers must decide upon with regards to neonatal units. These range from the ratio of different staff members to patients on the neonatal unit to the overall structure of neonatal healthcare services in a particular region. Many of these factors are studied in this

thesis.

Since 2003, neonatal units in England have been organised into networks, called managed clinical networks (MCNs) up until April 2013.[2] Within the managed clinical network framework, neonatal units are designated a particular care level based on their ability to provide a certain level of care to their patients. There are three key categories of care used in neonatal medicine. The British Association of Perinatal Medicine (BAPM) defines the general principle of these levels of care as follows (British Association of Perinatal Medicine, 2011):

I **Intensive Care**: This is care provided for babies who are the most unwell or unstable and have the greatest needs in relation to staff skills and staff to patient ratios.

II **High Dependency Care**: This is care provided for babies who require highly skilled staff but where the ratio of nurse to patient is less than intensive care.

III **Special Care**: Special care is provided for babies who require additional care delivered by the neonatal service but do not require either Intensive or High Dependency care.

The specific definitions in terms of the treatments that define each level of care are given in *Categories of Care* from the British Association of Perinatal Medicine (2011). There are three designations that a neonatal unit can receive: the largest, and most intensive unit, is a level three centre or network neonatal intensive care unit (NICU); the second is a level two or local neonatal unit; and, the least intensive is a level one or special care unit. Neonatal units are designated a level based upon their ability to provide each of the aforementioned levels of care. Each neonatal network, of which there were 23 in England up until April 2013, comprises one or two level three centres with

[2]The period of study that this thesis focusses on is 2006 to 2013 during which time the structure of the NHS and neonatal healthcare organisation was relatively stable. Following the Health and Social Care Act (2012), the organisation of the National Health Service, and of neonatal healthcare, changed in April 2013. As such data collected after this point are not used for the empirical analyses in this thesis. The background given here thus relates to the period 2006-13. However, the changes that occurred are also briefly detailed where appropriate. After April 2013, managed clinical networks were renamed operational delivery networks (ODNs) whose function is the same as MCNs but whose structure may differ.

a number of level two and level one units (the specific organisation of each network is given in Chapter 2). A dedicated neonatal transport team provides transfers between the units within a network, and occasionally between networks. Infants who require a higher level of care than is provided at the unit where they are admitted are transferred to an appropriate unit. Transfers back to the original unit can also be provided so that care may take place closer to the parental residence when possible (these are known as back transfers). Chapter 2 provides detailed definitions of key variables, the data that are used in this thesis, as well as specifics regarding contributing neonatal units, and an outline of their managed clinical networks.

Formal healthcare in England, including neonatal specialist healthcare, is predominantly provided by the National Health Service (NHS), which comprises a complex structure of various agencies involved in the commissioning, provision, and regulation of healthcare services. During the period of data collection for this thesis, 2006 to 2013, the structure of both the NHS in general and the specific organisation of neonatal specialist services remained relatively unchanged.[3] As a result of legislation in the late 1990s, between April 2002 and March 2013 independent organisations called Primary Care Trusts (PCTs) were responsible for commissioning primary, secondary, and community healthcare (Talbot-Smith and Pollock, 2006). The role of PCTs was to improve the health of their local healthcare community, to plan and secure the provision of services, and to integrate health and social care—as a result, PCTs spent around 80% of the total NHS budget (Department of Health, 2013b). For the period relevant to this thesis, there were 152 PCTs in England.[4] Generally, secondary and tertiary healthcare services such as neonatal specialist care were provided by individual or groups of hospitals arranged into NHS Trusts or NHS Foundation Trusts with which the PCTs contracted. Neonatal healthcare comprised approximately 1% of PCT budgets which amounted to a little under £1billion annually (this information is sourced from the Programme Budgeting Data (Department of Health, 2013a); see also Section 2.4 in

[3]Following the Health and Social Care Act (2012), the organisation of the NHS has changed. In April 2013, Primary Care Trusts were abolished and replaced with Clinical Commissioning Groups (CCGs). A full summary of changes can be found in *Health and Social Care Act: Fact Sheets* (Department of Health, 2012). The proceeding discussion focusses only on PCTs but applies for the most part to CCGs.

[4]Betweeen 2002 and 2005 there were 303 PCTs prior to a restructuring.

Chapter 2).

The challenges facing neonatal medicine have changed over the past two decades. The patient population has evolved in line with the reduction in the mortality rate among newborns, and, in particular, preterm infants. Between 1995 and 2006 the number of infants born between 22 and 26 weeks gestation admitted to neonatal care in England increased by 44% while over the same period the proportion of live births surviving to discharge at this gestation increased from 40% to 53% (Costeloe et al., 2012).[5] The change in the composition of the admitted population has led to the re-organisation of neonatal specialist healthcare in England and elsewhere.

In 2003, the Department of Health published the report of an expert working group convened "in order to provide advice on the most effective ways of caring for very sick or very premature newborn babies" (Department of Health, 2003). At this time, a small number of studies based in the United States had found evidence for a lower risk of mortality for very low birth weight infants born in hospitals with higher volume neonatal units, that is those units that deal with a greater number of patients or perform a greater number of procedures over a specific period of time (in particular Cifuentes et al. (2002) and Phibbs et al. (1996)).[6] On this basis, a key concern was to increase the access to high volume units for vulnerable infants, such as the VLBW or very preterm groups. One strategy to increase the proportion of births taking place in hospitals with high volume neonatal units involves centralising neonatal care by closing smaller neonatal units (centralisation is often also referred to as regionalisation, I use both terms interchangeably here). This strategy of centralisation has been advocated by a number of authors on the basis of evidence linking patient volume and outcomes (for example, Phibbs (2012) and Phibbs et al. (2007)), and has been enacted in countries such as Portugal and Finland with success (Binder et al., 2011; Neto, 2006). Regionalisation as a policy is by no means a recent phenomenon having been advocated since

[5]These are some of the most recently published statistics for the United Kingdom. However, with the advent of new neonatal data sources, in particular, the National Neonatal Research Database (NNRD), it is possible to update these statistics as frequently as each quarter as this thesis will demonstrate. The NNRD is described in detail in Chapter 2.

[6]Very low birth weight is a very commonly used classification for newborn infants. VLBW infants comprised 1.1% of all live births in 2012 (Office for National Statistics, 2014).

the 1960s for procedures requiring high skill, such as cardiac surgery, and was first discussed in the context of perinatal care in the 1970s (Philip, 2005; Usher, 1971). However, it was argued in the Department of Health report that centralisation may impair equity of access to neonatal healthcare services, particularly among those infants who may not benefit from high volume neonatal specialist healthcare, and for whom the burden of the long lengths of stay associated with neonatal care would be exacerbated by the increased distances families would have to travel (Department of Health, 2003). As such it was concluded that "in order to provide equity of access to care of the highest standard, which produces the optimal outcomes, neonatal care must be organised in a managed clinical network to ensure appropriate treatment" (Department of Health, 2003).

One of the primary aims of this thesis is to address whether the MCN arrangement in neonatal healthcare has produced the 'optimal outcomes' as specified by the Department of Health report (Department of Health, 2003). In particular, Chapter 4 examines the effects of neonatal unit volume and designation at the hospital of birth on adverse clinical outcomes for infants born in and admitted to neonatal care in England between 2009-11. This chapter provides the results from a retrospective, population based analysis of operational clinical data from the National Neonatal Research Database (NNRD). The NNRD extracts data from the electronic patient records of infants admitted to neonatal units in England. The NNRD now contains data from 165 neonatal units in England (95% of the total). This thesis summarises some of the first research projects conducted using these data and is the first research conducted into the effects of neonatal unit characteristics on patient health outcomes since the formation of MCNs in English neonatal healthcare. However, one of the issues with examining the relationship between neonatal unit characteristics and patient outcomes in a networked setting, and indeed in other healthcare settings, is that patients are transferred to units on the basis of their risk of mortality. The most severely ill infants can be transferred, either *in utero* or postnatally, to level three units, which tend also to be the units with the highest volumes of patients. As a result, the outcomes of infants admitted to these units may appear worse than if infants had been randomly assigned to a

neonatal unit. The analyses presented in Chapter 4 allow for the bias caused by these transfers, one of few studies of this nature to do so. Indeed, of hundreds of published studies investigating the link between hospital or procedure volume and patient outcomes, only a small number allow for the bias caused by this selective referral (Barker et al., 2011; Halm et al., 2002a). Chapter 4 finds that very preterm infants admitted to higher volume neonatal units, in terms of both patient numbers and caredays provided, at the hospital of birth are at a lower risk of mortality than their counterparts admitted to lower volume neonatal units at the hospital of birth. It also finds that the methods that do not account for the selective referral of infants between neonatal units generally underestimate the benefit of admission to a higher volume neonatal unit at the hospital of birth (results from this chapter have since been published elsewhere (Watson et al., 2014)).

The recipients of neonatal care represent a sizeable proportion of live births (9.0% in 2011, see Chapter 2 for recent statistics) and as such any intervention which impacts their health may have large, long-term consequences to overall population health. Therefore, it is imperative to understand the potential ramifications of organisational changes in neonatal care, in terms of staffing, resourcing, management, and cooperation, on both efficiency and equity outcomes. Alterations to health services that have a negative effect on outcomes, particularly in the case of neonatal care, may lead to an amplification of problems downstream with regards to morbidity and mortality. The health of an individual at birth has been shown to have consequences on later life educational, health, and labour market outcomes (Black et al., 2007). Thus, and as is argued in Chapter 8, these outcomes need to be considered over the course of a whole lifetime. Moreover, given the potentially large welfare implications that changes to neonatal care may have, there may be a case to increase provision of resources to neonatal healthcare, as the returns to each healthcare pound may be relatively large.

The topic of the health and social welfare returns for each pound spent on medical care is currently an important topic within healthcare research. New healthcare technologies or policies are evaluated in terms of the outcomes they would achieve versus their costs. In the United Kingdom, the agency responsible for producing guid-

ance on healthcare technology or policy efficacy or cost-effectiveness is the National Institute of Health and Care Excellence (NICE).[7] Commonly, technologies are evaluated for their cost-effectiveness: the ratio of incremental health outcomes achieved to incremental costs.[8] However, for each technology or policy that is recommended, resources currently being utilised within the healthcare system must be displaced to fund their adoption. The cost-effectiveness threshold should reflect the displacement implications of adoption decisions in terms of the health foregone which can be estimated by determining the effect of changes to current healthcare expenditure on the health outcomes of patients at the margin.

The aim of Chapter 5 is to estimate the returns to neonatal healthcare expenditure in terms of an incremental cost per statistical life saved or incremental cost per life year gained. Only a few studies have attempted to estimate the returns to healthcare spending currently being achieved in the National Health Service, the most recent and perhaps comprehensive example of which is Claxton et al. (2013). In this study, Claxton et al. (2013) utilise data on local area healthcare expenditure and health outcomes to estimate the cost-effectiveness threshold. However, in Chapter 5 I argue that the method and data used, in particular the use of aggregate expenditure and outcomes data, may not be suitable for the identification of the effect of interest. Thus, the work in Chapter 5 is important both methodologically and from a policy perspective. In the former case, it demonstrates how returns to medical spending can be estimated from routinely collected patient level data (in this case extracted from the NNRD), and, in the latter instance, it is shown that neonatal care may be cost-effective relative to other fields of healthcare. The analysis presented in Chapter 5 also explores how neonatal unit expenditure on healthcare is related to the various staffing levels and capital inputs reported by these neonatal units in a recent survey, the Unit Profile Survey, which is described in Chapter 2 and Chapter 5.

It may seem intuitive that increased healthcare expenditure, up to a point, should

[7]Prior to the Health and Social Care Act 2013, NICE was called the National Institute of Clinical Excellence; its new title is designed to reflect its expanded responsibility for guidance in areas of social care.

[8]This is distinguished from cost-benefit analysis where both costs and benefits are evaluated in the same unit which is typically present day monetary value.

translate into improved patient outcomes. It must also be considered that there may be a diminishing marginal benefit to healthcare expenditure so that increased expenditure may not result in improved health outcomes—this has been termed 'flat of the curve' medicine elsewhere (Fuchs, 2004). Nonetheless, while estimation of the returns to medical expenditure are important for the aforementioned reasons, improvement of neonatal unit productivity requires analysis of how expenditure translates into patient outcomes. The input mix, in terms of labour and capital, is likely to vary between neonatal units. As such, technical efficiency (the effectiveness with which inputs to healthcare translate into patient outcomes) may differ as well, there being various sources of (X-)inefficiencies between units. Much discussion has revolved around the optimal method of estimating productivity in the healthcare sector given the difficulty in quantifying what exactly 'hospital output' is as it is both multi-dimensional and involves health, which may be an unobservable construct. Moreover, there is further debate surrounding the correct specification of a hospital production function, the function used to describe how inputs translate into outputs.

The role of labour within the neonatal unit is key to understanding the processes underlying the empirical relationships previously discussed, such as the link between the volume of patients and their clinical outcomes or the relationship between expenditure and outcomes. The causal effect of unit volume on patient outcomes may be mediated through one of two mechanisms: economies of scale and learning by doing. The former mechanism is self-explanatory, i.e. that the long run average costs of a neonatal unit are lower the larger the unit is, which may be due to increased specialisation of labour, reduced cost of capital, or improved technical efficiency. One recent study has suggested that neonatal units are generally operating under increasing returns to scale in the United States (Leleu et al., 2012). The learning by doing mechanism is so named as it refers to the experience of the workforce.

Learning by doing has been an essential part of theories of productivity growth (Arrow, 1962; Lucas, 1988), and is an important source of competitive advantage for firms (Levitt et al., 2013; Yang and Borland, 1991). While neonatal units are not (effectively) in competition with one another for patients in the United Kingdom, the

units are organised into a networked system to take advantage of the improved ability of higher designation and volume neonatal units to treat the sickest infants. Moreover, given the importance with which the reorganisation of neonatal care was treated specifically in order to improve access to high volume and high designation neonatal units, this suggests that lower volume neonatal units may not be able to replicate the outputs of larger units. Larger units may be conceptualised as having a 'competitive advantage' over the lower volume neonatal units. Neonatal care, and particularly neonatal intensive care, is a complex process, and it is recognised that the first few hours of an infant's life are crucial in determining clinical outcomes. Indeed the term 'golden hour' is used more and more frequently within neonatal units to refer to the first sixty minutes of an infant's life (Doyle and Bradshaw, 2012). The golden hour requires a number of team-orientated and task-based protocols to stabilise an infant through thermoregulation, antibiotic administration, and establishment of appropriate nutrition among other tasks (Doyle and Bradshaw, 2012). These tasks are complex and require skill on the part of the team administering them; the role of learning by doing in improving neonatal outcomes becomes clear when the number of complex tasks that need to be completed efficiently and effectively are considered. The tacit knowledge involved in neonatal intensive care obtained through learning by doing is unlikely to be transferable between units, since even if skilled clinicians were transferred from one unit to another, skills can degrade over time through a process of forgetting or neglect (for example, Gaynor et al. (2005); Huesch (2009); Sfekas (2009)). Moreover, the specific skills of the clinical staff in neonatal units may be both inimitable and non-substitutable. In this sense, the specific human capital accumulated by large neonatal units is what constitutes their advantage.[9] However, neonatal intensive care requires a complex interaction of clinical skills with various types of capital. Economies of scale are likely to complement the advantage of learning by doing since the unit will have at its disposal greater resources as well as a superior ability to deploy those resources

[9]Specific human capital is as opposed to general human capital which Kim and Mohtadi (1992) distinguish as 'The former is a stock of specialized knowledge and skills that improves worker productivity in a given production activity; the latter is a stock of general knowledge that renders the workers more adaptable to a variety of activities.'

in the provision of neonatal healthcare. Indeed, both economies of scale and learning by doing have been observed operating at the same time in other areas of healthcare (Gaynor et al., 2005; Huesch, 2009; Sfekas, 2009).

Taken together, these points suggest that the networked system in England may not be able to replicate the outcomes of a fully centralised system because, firstly, during the golden hour the skills of the staff at the hospital of birth may not be as developed in lower volume or designation neonatal units regardless of subsequent transfers; and secondly, the skills of the staff at lower volume and designation neonatal units may not be able to be improved simply through training at higher volume or designation neonatal units or by use of other techniques or forms of labour. This thesis provides analyses that shed light on these questions and aid healthcare policy makers.

Whether or not a networked system is able to improve patient outcomes, the question remains as to whether the advantage of high volume or designation neonatal units may be replicated in other ways. The factors which confer an advantage to firms operating in a 'standard' market generally do not apply in a centralised, public healthcare system. For example, in a rapidly changing technological environment, such as that characterised by the healthcare sector, technological innovations can give a firm a competitive advantage. However, in a public healthcare system, provided technologies meet predefined cost-effectiveness or cost-benefit criteria, these technologies may be available to all hospitals. Moreover, the processes by which labour and capital inputs translate into health outputs within the neonatal units are generally clear so that processes can be easily replicated between units; and, there are no 'switching costs' faced by patients beyond the different travel times required to access different hospitals. The remaining explanations for the advantage conferred by large neonatal units then revolve around the specific capabilities of those units. However, since these capabilities are potentially non-transferable, this may suggest that the advantage of large neonatal units cannot be eliminated. It should also be considered that as newer technologies are developed that reduce the mortality rate, the complexity of the procedures utilised in neonatal intensive care may increase, which should further increase the benefit of birth in a hospital with a high volume or designation neonatal unit. This leads to the question

of the appropriate outcomes used in the evaluation of neonatal healthcare technology. Evaluation of benefits on the basis of the mortality rate may lead to the adoption of technologies that save infants at the margin for the risk of mortality for whom the disability free survival rate is very low. Indeed, a more appropriate measure of benefit may be the overall life years gained weighted by their quality, as is generally the case in other areas of healthcare (Culyer, 2010). However, there currently do not exist measures of validated measures of health related quality of life for infants being treated on neonatal intensive care units (Boss et al., 2012), nor are there reliable data relating to the long term outcomes of these infants. Chapter 8 considers how benefits for neonates ought to be considered and how the empirical evidence presented in this thesis ought to be assimilated in healthcare policy. Nonetheless, there is, as yet, no evidence of the effect of unit volume on patient outcomes in a networked system, in neonatal or any other form of formal healthcare.

With regards to the labour inputs to neonatal units, there exist recommended nurse to patient ratios for various intensities of neonatal healthcare. For example, the British Association of Perinatal Medicine (BAPM) recommends a one to one nurse to patient ratio for neonatal intensive care (British Association of Perinatal Medicine, 2011). Maintaining these ratios may be essential for the effective utilisation of the available capital inputs. Part of the nurses role in intensive care involves, for example, the monitoring of blood oxygen levels and core temperature, which require constant attention, as well as human input such as touch (Boxwell, 2010). Many of the roles performed by nursing labour cannot be substituted for with capital inputs. Neonatal intensive care may therefore benefit from both increased labour as well as increased specific human capital embodied in that labour. This is particularly true if neonatal units are not capital constrained, which is likely to be the case in England since many neonatal units close cots (i.e. do not permit new admissions into these cots) due to a lack of available labour, rather than there not being enough cots to occupy the available nursing staff (Parmanum et al., 2000). However, the nurse to patient ratios recommended by the BAPM are based on time use studies conducted in the early 1990s (Northern Neonatal Network, 1993; Williams et al., 1993), since which time the role of nurses has changed

dramatically given technological changes and differences in the patient population.

Chapter 6 aims to determine whether increasing the proportion of infants receiving intensive care who receive one to one nursing improves the outcomes of those infants. As with the previously described empirical chapters, there is the issue that the 'treatment' under examination, which is one to one nursing in this case, is not randomly assigned to patients. Infants who are at higher risk of adverse clinical outcomes, such as mortality, are more likely to receive more intensive nursing than other infants. In this chapter, I use a novel strategy for examining the effect of one to one nursing among neonatal intensive care patients, that involves the aggregation of data from the NNRD to the neonatal unit level for each month over a period of 60 months. I develop a simple theoretical model of the allocation of nursing labour to different tasks, which is used to inform the empirical specification used in Chapter 6. To preview the findings from this chapter, it is found that a one percentage point increase in the proportion of intensive care days with one to one nursing leads to a reduction in the mortality rate of 0.14 percentage points (compared to a mortality rate of 4.5% for the infants in the sample considered in this chapter).

Neonatal specialist care units in hospitals respond to the changing needs of the birth cohort. In recent years, the proportion of live births admitted to neonatal care has increased from 7.6% of births in 2007 to 9.0% in 2011 (see Chapter 2 and Chapter 7). One possible interpretation of this increase is that the average health of live births in this period has deteriorated.[10] Understanding the causes of this increase is important both to ensure there is adequate provision of neonatal care in the future, but also to guide future policy aimed at improving health, both intra- and inter-generationally.

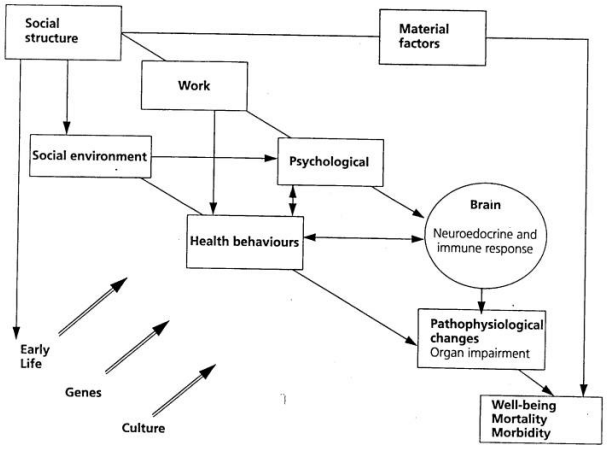
The genesis and existence of health inequalities in the population and the relationship between income and health inequalities are the subject of much debate and enquiry (Marmot, 2004). The processes by which health is influenced by social factors are complex and interlinked; Figure 1.2 shows a model of the social influences on health. It is possible to simplify the model somewhat by dividing the causes into ma-

[10]Another interpretation being that the threshold for admission to neonatal healthcare has changed in this period, Chapter 7 considers both of these interpretations.

terial and relative causes. Material causes may include access to medical care, ability to consume a healthy diet, health related behaviours; relative causes are psychosocial stressors, including work environment and social integration (Marmot, 2004). However, there is a third pathway that is becoming more prominent where it is hypothesised that the nine months *in utero* are perhaps the most important in shaping an individual's future health trajectory, called the Fetal Origins Hypothesis (Almond and Currie, 2011). Shocks and inputs at this time are crucial to future health outcomes. Importantly for this thesis, a growing body of evidence supports the Fetal Origins Hypothesis. Almond and Mazumder (2011) find that prenatal exposure to Ramadan among infants in Michigan, United States leads to lower birth weight and that Muslims in Uganda and Iraq are 20 percent more likely to be disabled as adults if they were exposed to Ramadan *in utero*. Meanwhile, Black et al. (2007) find that a 10 percent increase in birth weight, on average, increases the probability of high school graduation by 1.2 percent, IQ (of men) by 1.2 percent, earnings by 0.9 percent, and height by 0.3 percent.[11] Given this body of evidence, maternal and neonatal healthcare grows in its importance, since healthcare interventions here may have a crucial effect in reducing health inequalities. As has been previously documented, admissions to neonatal care come disproportionately from more socio-economically deprived households (see for example, in the UK, Smith et al. (2009); Smith and Hall (2011)), reflecting a commonly observed trend in practically all areas of healthcare. In this way the conceptualisation of the healthcare system as a redistributive institution becomes clear. The healthcare system acts to mitigate medical conditions or poor health that are, in many cases, generated by other social institutions. In this institutional context, when considering the role of healthcare and policies relating to it, we are interested not only in the distribution of health outcomes, but how these outcomes are produced. Healthcare is generally reactive to changes in population health but may be more effective if capacity could be effectively managed to reflect need, or, preferably, upstream causes of ill health and disease may be targeted to reduce the burden on healthcare systems.

[11] A comprehensive review of recent developments in the area of the Fetal Origins Hypothesis can be found in Almond and Currie (2011).

Fig. 1.2 Model of pathways of social influences on health



Source: Marmot (2004).

Factors underlying the increase in the rate of admissions to neonatal healthcare between 2006 and 2011 are also investigated in this thesis. Specifically, Chapter 7 aims to estimate the effect that changes to local economic conditions have on population infant health at birth; the unemployment rate was observed to increase over the same period as the observed increase in admissions to neonatal care and is well correlated with local economic conditions. Previous studies have sought to determine a relationship between local economic conditions at conception and infant health at birth (Dehejia and Lleras-Muney, 2004). However, birth weight may not be a complete measure of infant health at birth. While birth weight is correlated with underlying health status at birth, it only captures one part of the distribution of health, and may not lie on the causal pathway that generates the observed health outcomes (Wilcox, 2001). As a result, Chapter 7 uses a novel measure of infant health at birth—the proportion of live births admitted to neonatal healthcare—it also uses birth weight as an outcome but focusses on the rate of admissions to neonatal care as a more complete measure of the health of the cohort of live births. Within area effects are estimated using national birth data along with detailed operational clinical data from the NNRD. This chapter finds evidence that increases in local unemployment at the time of conception leads to increases in the proportion of live births admitted to neonatal specialist healthcare. It also finds that this effect is not mediated by changes to birth weight, and that this effect

is only evident in the most socio-economically deprived areas. These results provide new evidence of the effects on infant health of changes to the local unemployment rate and suggest that increases to local unemployment may exacerbate socio-economically determined health inequalities.

The opportunities to improve neonatal healthcare in the future are rich and varied. In the past, much evidence regarding best practice in neonatal healthcare has relied on reported clinical experience or single-site evidence (Philip, 2005). Presently, with the advent of electronic reporting systems, and improvements to information technology, large data sources are being curated enabling expansive multi-site investigations to take place. The first such example of a collaborative initiative was the Vermont-Oxford Network (VON) which began with 34 neonatal intensive care units in 1989 and is now comprised of almost 1000 centres (Vermont-Oxford Network, 2014). In the United Kingdom, a similar initiative, the NNRD was established in 2006 at Imperial College, London (these data are described in detail in Chapter 2). As previously mentioned, the NNRD data are used throughout this thesis, and the empirical analyses presented here represent some of the first using this data. This thesis thus provides a number of important contributions. The results from the econometric analyses presented in Chapters 4 to 7 provide significant evidence for neonatal healthcare policy makers aimed at producing optimal health outcomes from neonatal healthcare organisation. The methodology as well as the choice of data sources and variables represent a novel contribution to a number of important research questions, such as the returns to healthcare expenditure. And finally, this thesis demonstrates the power of large, national, collaborative data sources such as the NNRD.

The rest of this thesis is structured as follows. The next chapter, Chapter 2, details the data sources employed in this thesis and explores key variables. Chapter 3 reviews the literature on the relationship between neonatal unit characteristics and infant clinical outcomes. Chapter 4 estimates the effect of neonatal unit volume and designation on infant mortality and a range of morbidity outcomes. Chapter 5 explores the effect of neonatal unit healthcare expenditure on the risk of mortality. Chapter 6 estimates the effect of the one to one nursing rate on the risk of mortality for infants receiving inten-

sive care on neonatal units. And, in the final empirical chapter, Chapter 7 investigates the effect of local economic conditions on infant health at birth. Chapter 8 discusses the normative issues surrounding the use of the empirical evidence presented in this thesis in the formulation of neonatal healthcare policy. Finally, Chapter 9 concludes.

Chapter 2

Data Sources and Key Variables

2.1 Introduction

A variety of data sources are utilised in this thesis which are summarised in this chapter. Key variables are also defined and described. Four key data sources are employed: (i) the National Neonatal Research Database (NNRD) which provides patient level, clinical and process data, which is introduced in Section 2.2; (ii) the Office for National Statistics (ONS) which provides population demographic and socio-economic data, these data are summarised in Section 2.3; (iii) the National Health Service (NHS) which provides data on healthcare provider costs and healthcare authority expenditure, which are discussed in Section 2.4; and, (iv) the Unit Profile Survey (UPS) 2011, which contains data on neonatal unit staffing and resourcing and is detailed in Section 2.5. The NNRD data are used in all the empirical chapters in this thesis, both at the individual level (in Chapter 4 and Chapter 5) and aggregated to neonatal unit and local area levels (in Chapter 6 and Chapter 7). This chapter also introduces and defines the key variables used in the empirical analyses, including clinical variables such as gestational age and birth weight, as well as economic variables such as the unemployment rate. Certain variables, such as one to one nursing, where they are key to certain analyses are presented in their relevant chapters. However, the contents of this chapter should inform all of the emprical analyses presented in this thesis.

2.2 The National Neonatal Research Database

The National Neonatal Research Database (NNRD) was created by the Neonatal Data Analysis Unit (NDAU), a research unit based at Imperial College, London, and was established in 2006, using the individual, electronic patient records of infants treated within neonatal units in England. The NDAU holds national research ethics committee approval to create this database (reference REC 10/H0803/151) as well as the permission from the Caldicott Guardians of each NHS Trust.[1] The data are psuedo-anonymised by removing patient and maternal identifiers and encrypting the NHS number of each infant. The data include a vast range of variables; including static descriptive variables captured once per baby, such as birth weight and gestational age at birth; episodically, such as episodes of infection and other clinical outcomes; and daily items such as treatments and procedures as well as level of neonatal care.

The electronic patient records from which the NNRD are extracted are completed by a variety of staff at the contributing institutions. This may lead to discrepancies in data quality and missing data; this is discussed in the following section, Section 2.2.1. Often it is the nurse assigned to a particular infant on the night shift that completes the record; the software that manages these electronic patient records is called Badgernet or Badger 3 depending on the version utilised. The process of pseudo-anonymisation is carried out by a company called Clevermed which acts on behalf of NDAU. Initially, the Badger software and the centralised, electronic patient records system was adopted by the South East Neonatal Network (SEND) and surrounding neonatal units (which covered much of London, the South East and East of England). This is reflected in Table 2.1 which lists the number of units per region contributing to the NNRD; in 2006 neonatal units in the South East and East of England comprised 51.5% of the neonatal units in the database, by 2013 they only comprised 33.9% (row (7), Table 2.1).

The NNRD is the primary data source for much of the analyses presented in this thesis. In order to use the NNRD data for this purpose, contributing neonatal units

[1]The Caldicott Guardians are the designated individuals responsible for providing permission for use of data within each neonatal unit.

were offered the option to opt-out of this research. Overall, permission was obtained by 2011 to use data from 165 neonatal units in England (these units and their clinical leads are listed in the Acknowledgements).

2.2.1 Missing Data

A detailed overview of the data and summary statistics of the neonatal patient population are provided in this section. However, it should be noted that the data contained within the NNRD are extracted from electronic patient records that facilitate care at neonatal units. They are not explicitly collected to conduct academic research, rather, the data are primarily collected for audit purposes in the course of health care provision and are later extracted. Data collection is therefore carried out separately at each contributing institution by a variety of healthcare professionals. Thus, data quality may vary both over time and between units, as there are improvements in data capture, and between fields, where more commonly used fields are recorded more accurately and contain less missing data. Table 2.2 shows the rates of missing data for key variables. It is clear from this table that certain variables have fewer missing data than others. For example, gestational age at birth is completed for >99.9% infants in all years, compared to administration of antenatal steroids to an infant's mother which, in the most recent year, was missing for 8.3% of infants. It is also noted that there is an improvement in quality in 2010 in the proportion of missing data, this is due to a change in the method used to process and collate the data used by Clevermed and NDAU at this time.

It is important to differentiate between the various types of missingness since this may have repercussions on the empirical analyses presented later. Rubin (1976) defined a taxonomy of missingness based upon the mechanism generating the missing data. If this mechanism does not depend on the values (or potential values) of the variables in the design, then the data are missing completely at random (MCAR). In the case of the electronic patient records from which the NNRD is extracted, MCAR data may occur if random mistakes are made in data input. Data are missing at random

Table 2.1 Descriptive statistics of the NNRD data

	Variable	2006	2007	2008	2009	2010	2011	2012	2013
					A: All Neonatal Units				
(1)	N	96	128	146	150	164	165	162	159
(2)	East Midlands	4 (4.2)	8 (6.2)	12 (8.2)	12 (8)	12 (7.3)	13 (7.9)	13 (8.0)	13 (8.2)
(3)	East of England	21 (21.9)	21 (16.4)	24 (16.4)	24 (16)	22 (13.4)	23 (13.9)	22 (13.6)	22 (13.8)
(4)	London	31 (32.3)	30 (23.4)	34 (23.3)	35 (23.3)	32 (19.5)	33 (20)	33 (20.4)	32 (20.1)
(5)	North East	3 (3.1)	5 (3.9)	14 (9.6)	15 (10)	13 (7.9)	14 (8.5)	14 (8.6)	14 (8.8)
(6)	North West	13 (13.5)	26 (20.3)	29 (19.9)	29 (19.3)	27 (16.5)	28 (17.0)	26 (16.0)	25 (15.7)
(7)	South East	28 (29.2)	31 (24.2)	34 (23.3)	34 (22.7)	32 (19.5)	33 (20)	33 (20.4)	32 (20.1)
(8)	South West	11 (11.5)	17 (13.3)	21 (14.4)	21 (14)	19 (11.6)	20 (12.1)	20 (12.3)	20 (12.6)
(9)	West Midlands	6 (6.2)	10 (7.8)	16 (11.0)	19 (12.7)	21 (12.8)	22 (13.3)	22 (13.6)	22 (13.8)
(10)	Yorkshire and the Humber	3 (3.1)	4 (3.1)	10 (6.8)	9 (6)	18 (11.0)	19 (11.5)	19 (11.7)	19 (11.9)
(11)	Level 3	26	36	42	43	45	44	43	43
(12)	Level 2	50	66	71	72	74	75	74	73
(13)	Level 1	20	26	31	34	45	46	45	43
(14)	Admissions[a]	171.0 (169.8)	273.9 (190.4)	286.2 (232.4)	333.5 (268.5)	339.8 (275.7)	381.9 (307.3)	422.0 (339.8)	431.0 (337.9)
(15)	Caredays[b]	2,795.7 (3,146.9)	4,090.2 (3,003.0)	4,479.1 (3,252.4)	5,073.9 (3,451.4)	4,630.2 (2,925.4)	5,241.5 (3,237.1)	5,400.3 (3,405.0)	5,189.3 (3,281.4)
(16)	Birth weight (g)	2,552.4 (963.1)	2,634.7 (940.0)	2,657.1 (949.3)	2,706.5 (935.3)	2,771.5 (911.4)	2,807.0 (901.3)	2,849.2 (888.1)	2,876.1 (874.9)
(17)	Gestational age (wks)	35.6 (4.2)	36.0 (4.0)	36.0 (4.1)	36.3 (3.9)	36.6 (3.8)	36.8 (3.7)	36.9 (3.6)	37.0 (3.5)
(18)	Mortality (%)	1.9	1.7	1.9	1.5	1.9	1.8	1.5	1.3
					B: Balanced Panel				
(19)	Admissions[a]	171.0 (169.8)	321.4 (190.3)	350.8 (241.6)	397.6 (290.5)	390.8 (320.1)	423.4 (353.0)	462.2 (392.3)	458.6 (391.2)
(20)	Caredays[b]	2,795.7 (3,146.9)	4,815.6 (2,972.6)	5,399.1 (3,056.8)	5,782.8 (3,386.4)	5,184.0 (2,851.8)	5,639 (3,151.3)	5,707.9 (3,284.7)	5,371.1 (3,223.3)
(21)	Birth weight (g)	2,552.4 (963.1)	2,651.1 (935.8)	2,690.6 (944.0)	2,742.3 (930.6)	2,811.5 (905.3)	2,853.0 (895.6)	2,894.9 (879.7)	2,920.0 (861.8)
(22)	Gestational age (wks)	35.6 (4.2)	36.0 (4.0)	36.2 (4.0)	36.4 (3.9)	36.7 (3.7)	37.0 (3.6)	37.1 (3.5)	37.2 (3.4)
(23)	Mortality (%)	1.9	1.6	1.8	1.5	1.8	1.6	1.4	1.2

[a] Average annual number of unique infants admitted to neonatal units contributing to the NNRD
[b] Average annual number of care days provided neonatal units contributing to the NNRD
Panel A shows descriptive statistics of all units contributing data to the NNRD. Panel B shows descriptive statistics only from those units contributing data in every year of the NNRD.
Values are mean (sd) unless otherwise stated.

(MAR) if the probability of data being missing for a variable is not a function of that variable conditional on some other variable in the design. This definition of MAR is the same as used by Allison (2001) who states that data are MAR if 'the probability of missing data on Y is unrelated to the value of Y, after controlling for other variables in the analysis.' For example, if the mortality outcomes for very preterm infants were less likely to be missing than for their non-very preterm counterparts, but the risk of mortality was not related to the missingness of the outcome data, then these data would be MAR. In the case of MCAR and MAR, the missingness is referred to as ignorable, since this missingness should not lead to bias in estimators of statistical models. However, if neither of the above conditions are met then the data are said to be missing not at random (MNAR). MNAR is non-ignorable since this will cause bias. Nonetheless, the mechanism generating missing data may be modelled, and certain estimators, such as that described by Heckman (1979), can be used which are generally consistent in the case of data that are MNAR. Indeed, a similar mechanism is used in the analyses presented in Chapter 7 to account for missing data at the neonatal unit level where the data suggest that there is a difference, at the local area level, between areas for which data are available are those for which they are not (see Table 7.1 in Chapter 7).

In the rest of this chapter, where data are missing, the missingness mechanism will be discussed. What is immediately noticeable from Table 2.2 is that the proportion of data which is missing reduces year on year. This suggests that the quality of data input and capture is improving over time.

2.2.2 Neonatal Unit Characteristics

The number of neonatal units contributing data to the NNRD has steadily increased since the inception of the NNRD.[2] Row (1) of Table 2.1 shows the total number of units contributing data in each year between 2006 and 2013. Figure 2.1 shows the locations and networks of contributing neonatal units. Since 2006, the number of contributing units has increased from 96 up to over 160 from 2010 onwards. In some years the

[2]The total number of units in England over the same time period has changed year on year due to closures and mergers but is approximately 170 over the whole period.

Table 2.2 Percentages of missing data by variable and year for selected variables in the NNRD

Variable	2006	2007	2008	2009	2010	2011	2012	2013
Gestational age (wks)	0.09	0.03	0.01	0.02	0.04	0.02	0.02	0.01
Birth weight (g)	1.52	1.35	1.14	1.10	0.03	0.03	0.02	0.01
Steroids[a]	30.31	19.03	19.12	19.18	10.56	9.44	9.25	8.22
Sex (% male)	1.75	1.40	1.22	1.16	0.04	0.03	0.21	0.03
Multiple birth (%)	1.64	3.15	1.52	1.11	0.07	0.03	0.03	0.01
Birth year	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Apgar (1 min)[b]	14.29	18.69	18.09	19.81	20.46	20.57	21.78	21.78
Apgar (5 min)[b]	15.11	19.44	18.86	20.43	20.90	20.95	22.14	22.20
Mum residence[c]	2.71	5.32	5.33	13.39	5.60	5.20	6.70	8.51
Mode of del.[d]	29.95	29.65	28.06	28.81	20.68	21.36	21.95	20.94
Place of birth	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Final outcome[e]	7.21	4.42	3.96	3.61	0.25	0.08	0.45	0.45

[1] Values are %

[a] Full or partial course of antenatal steroids administered to the mother prior to birth. See Section 2.2.3.

[b] APGAR score is a method to quickly assess the health of a newborn baby at birth (Casey et al., 2001).

[c] The location of the maternal residence is recorded. See Section 2.2.4.

[d] The mode of delivery is recorded.

[e] Final outcome refers to whether it is recorded whether an infant died or was discharged from neonatal healthcare.

number of units decreases; this is due to unit mergers and closures as opposed to units withdrawing permission to use their data.

Unit Designation and Volume

The annual volume of care a neonatal unit provides along with its designation will form a key part of the empirical analyses in this thesis. Unit designation is assigned by BAPM; BAPM provide the following official definitions:

1 **Special care unit**: Units provide Special Care but do not aim to provide any continuing High Dependency or Intensive Care. This term includes units with or without resident medical staff.

2 **Local neonatal unit**: Units provide High Dependency Care and some short-term Intensive Care as agreed within the network.

3 **Network neonatal intensive care unit**: Units provide the whole range of medical neonatal care but not necessarily all specialist services such as neonatal surgery(British Association of Perinatal Medicine, 2010).

Unit designation is therefore based upon a unit's activity and ability to provide different levels of care (defined in Chapter 1) as well as its role within its own managed clinical network (now renamed operational delivery networks). There may therefore be some discrepancy between networks with regards to designation. Indeed, some units designated local neonatal units provide a greater volume of care than network neonatal intensive care units. The volume of a neonatal unit refers to the total number of patients, care days, or procedures performed by the unit over a specific period of time. In this thesis, two principal measures of volume are used: the annual number of (unique) admissions and the annual number of care days provided.[3] In each case the measure may be further delineated by the recipient of the care or the intensity of care provided. For example, the primary measure of volume considered in Chapter 4 is the annual number of very preterm admissions.

[3]Some infants may be transferred away from a unit and then readmitted. 'Unique' admissions refers to only counting each individual once per unit regardless of the number of times the individual is admitted to the unit.

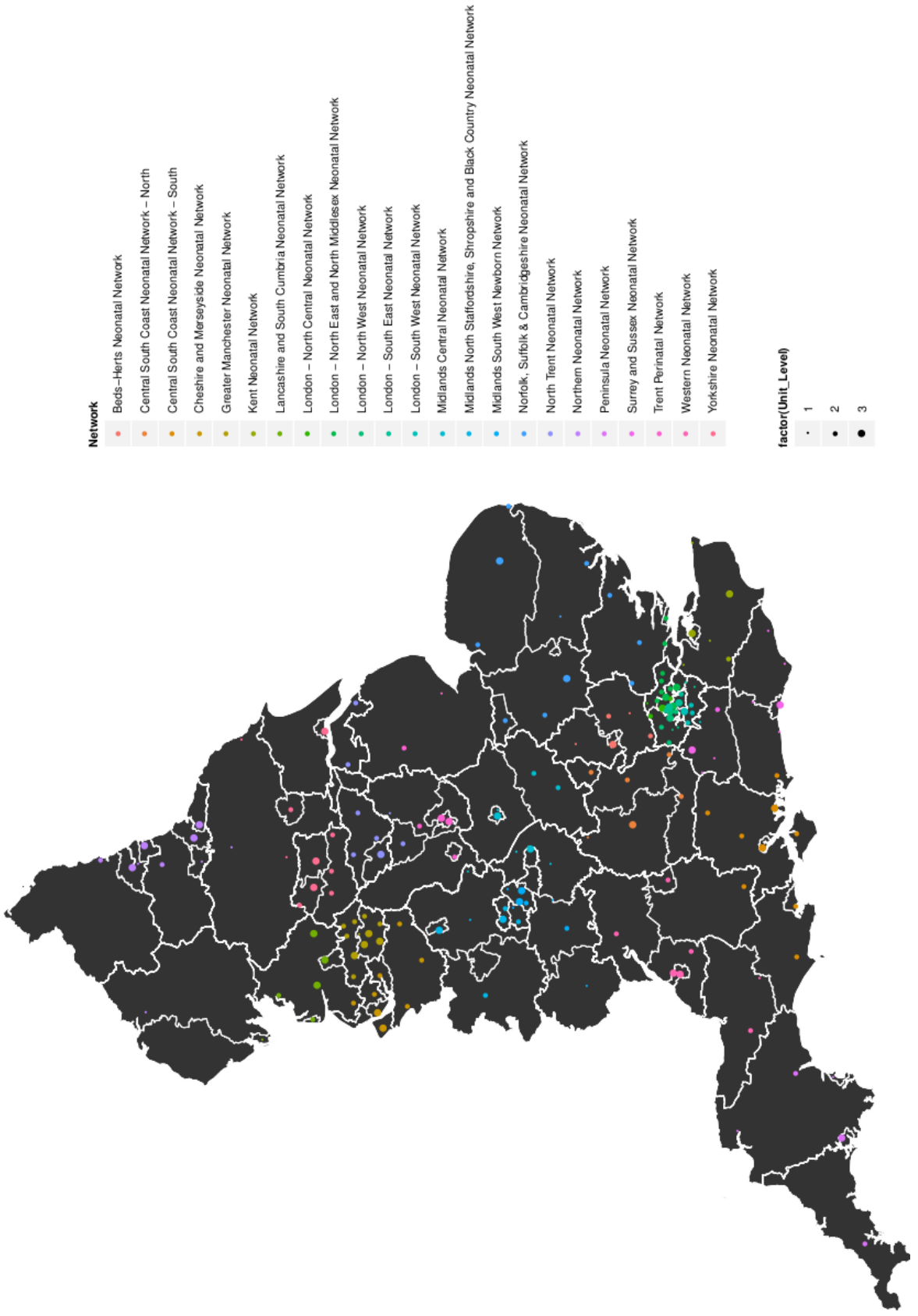


Fig. 2.1 Map of neonatal units in England by neonatal network

The subject of Chapter 4 is the effect that unit volume and designation has on infant clinical outcomes. As expected, the total number of care days provided by a unit varies with its designation, with level three units providing the highest volume of care (Figure 2.2).[4] What is also clear from Table 2.1 is that the average volume of care per neonatal unit has increased over the course of the panel, both when measured in terms of the number of admissions and the number of care days (rows (14) and (15) of Table 2.1). This may be observed in the data if the new units contributing each year were larger than the sample average; however, the lower panel of Table 2.1 is based on data from units contributing data in each of the eight years of the NNRD—these show a very similar pattern of increased volume of care (rows (19) and (20) of Table 2.1). Changes to local economic conditions may be a factor that has contributed to this increase, through mechanisms discussed in Chapter 1; this is the subject of the analysis in Chapter 7.

2.2.3 Clinical Variables

Much of the empirical work in this thesis involves the estimation of models of neonatal mortality and morbidity outcomes. Certain variables feature almost ubiquitously in individual level models of neonatal mortality and health in the literature (see Chapter 3 and the literature review by Medlock et al. (2011)). These baby level variables are gestational age at birth, birth-weight, sex, whether the infant was a multiple birth, and whether the mother received antenatal steroids. These are examined in this chapter.

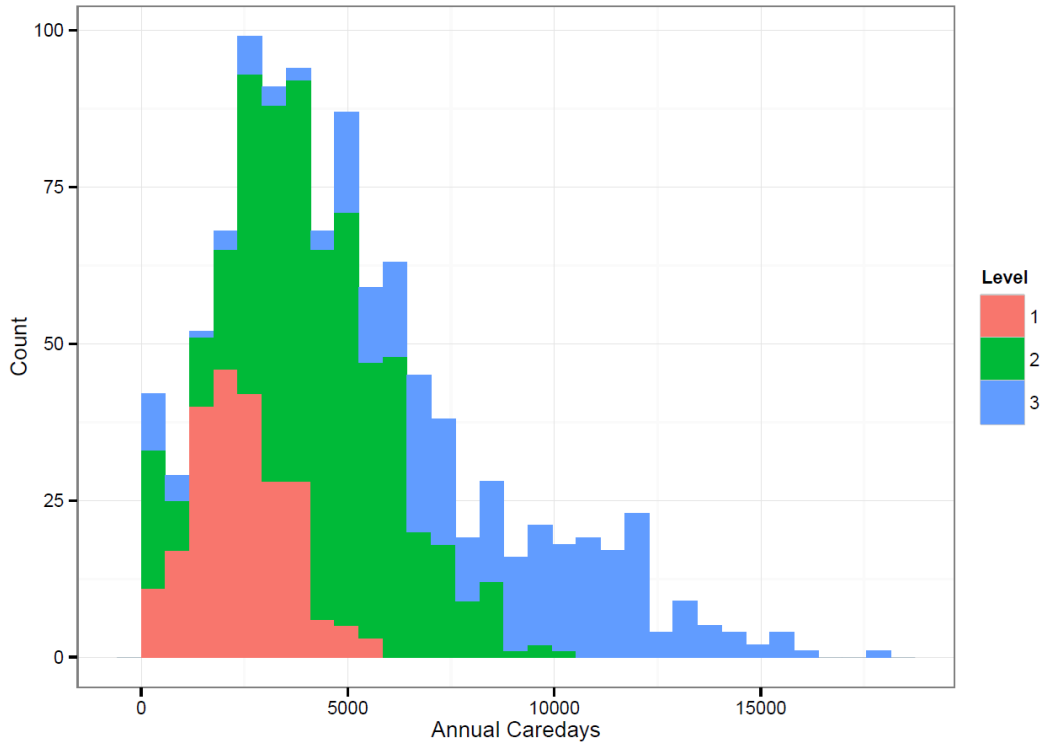
Gestational Age

Gestational age is the measure of the length of pregnancy and is calculated using obstetric ultrasonography in these data.[5] The gestational age of an infant is notated in this thesis as $weeks^{+days}$ as is typical in the clinical literature. A pregnancy is con-

[4] This figure is based on raw data from the NNRD. There are clearly a number of level two and three units with small numbers of admissions. These small numbers are, in many cases, due to units not contributing for the entire year.

[5] Gestational age may also be calculated as the time since the last menstrual period or the duration since fertilisation (plus 14 days).

Fig. 2.2 The distribution of annual care days provided by unit designation



Annual care days are the total number of care days provided to all infants at any level of care and recorded in the NNRD.

sidered 'at term' when the gestation has lasted at least 37^{+0} weeks but not more than 41^{+6} weeks. Births prior to this are referred to as 'preterm' and births after this period are 'post-term'. Each of these categories can be further delineated, these definitions are shown in Table 2.3 along with the percentage of admissions that each category represents. Figure 2.3a shows this distribution graphically.

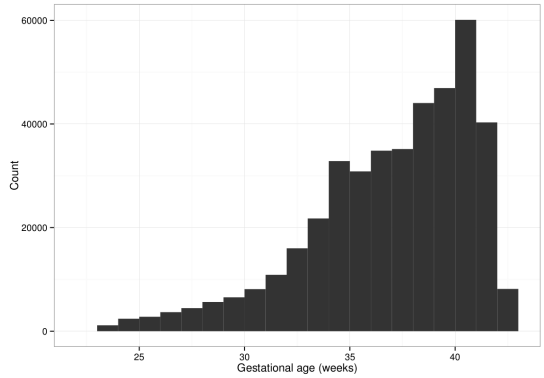
The most well represented group of admissions are moderately preterm infants (see Table 2.3 for definitions of gestational age groups), representing 32.9% of admissions. However, it is also clear from Table 2.3 that the burden of mortality is greatest among the very and extremely preterm groups of infants with the latter group comprising 48.3% of all neonatal unit deaths despite representing only 3.2% of the patient population. Figure 2.3b shows graphically the relationship between gestational age and risk of mortality; there is clearly a much higher risk amongst preterm infants. By around 34 weeks gestation, the risk of mortality falls to almost zero. The earliest term infants generally admitted to neonatal units are born at 23 weeks gestation—this is typically referred to as the limit of viability since disability free survival below this point is very low (Allen et al., 1993; Doyle, 2001). Survival of infants born at 22 weeks and 23 weeks gestation is 3% and 26% respectively, with corresponding disability free survival rates 1% and 13% (Costeloe et al., 2012). A similar relationship is observed between gestational age and the other adverse clinical outcomes considered in this study; these are discussed in Section 2.2.3.

The definitions of gestational age categories may seem extensive, however, recent research has shown clinical outcomes to vary by gestational age week (see for example Doyle (2001), Boyle et al. (2012)). Indeed, the definitions of term birth above have been recently recommended by the American College of Obstetricians and Gynaecologists (ACOG) in order to deter any avoidable birth prior to 39 weeks (American College of Obstetricians and Gynaecologists, 2013).

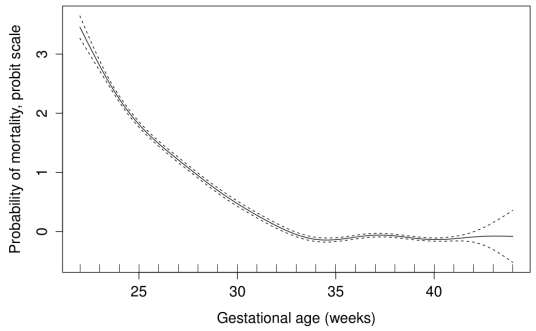
Birth weight

Birth-weight, like gestational age, is strongly correlated with adverse clinical outcomes such as mortality and is strongly correlated with gestational age (Figure 2.3c). For

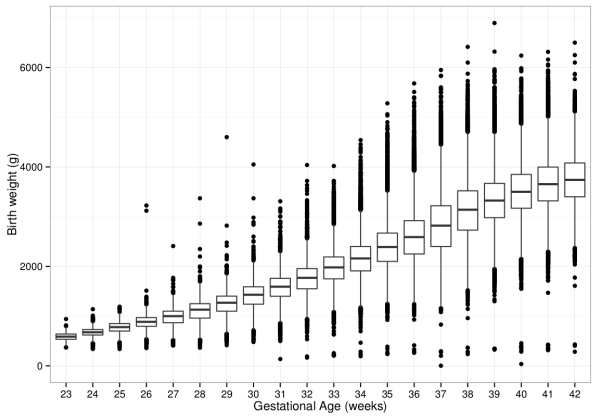
Fig. 2.3 Gestational age



(a) Histogram showing gestational ages of infants admitted to neonatal units



(b) Estimated function showing the relationship between gestational age risk of mortality.[a]



(c) Box plot of birth weight by gestational age week

Source: NNRD, 2006-13. [a]Estimated using a probit model with a cubic spline term for gestational age and dependent variable equal to one if the infant died in hospital and zero otherwise.

Table 2.3 Definitions of gestational age

Category	Definition	Percentage of admissions[a]	Percentage of deaths[b]	Median (IQ range) length of stay (days)[c]
Extremely preterm	$< 28^{+0}$	3.19	48.28	88.2 (71.2, 109.7)
Very preterm	$28^{+0} - 31^{+6}$	7.28	15.30	40.8 (30.5, 54.2)
Moderately preterm	$32^{+0} - 36^{+6}$	32.94	13.62	11.0 (5.0, 17.5)
Early term	$37^{+0} - 38^{+6}$	19.10	8.74	3.4 (1.4, 6.7)
Full term	$39^{+0} - 40^{+6}$	25.71	9.56	2.7 (1.2, 5.6)
Late term	$41^{+0} - 41^{+6}$	9.73	3.84	2.9 (1.2, 5.6)
Post-term	$\geq 42^{+0}$	2.05	0.62	2.9 (1.3, 5.5)

[a] Percentage of all admissions to neonatal units represented by each category
[b] Percentage of all deaths that occurred on neonatal units represented by deaths of infants in each category.
[c] Length of stay among infants surviving to discharge only.
[1] IQ range = interquartile range
[2] Source: data from the NNRD, 2006-13.

this reason, it is typical to use either gestational age (in the way described above) or birth-weight to classify infants for empirical research or clinical practice. For example, many studies based in the United States often classify infants as either low birth weight (LBW; $\leq 2,500g$) or very low birth weight (VLBW; $\leq 1,500g$) (for example Baker and Phibbs (2002); Phibbs et al. (2007)). Indeed, many treatments are assigned to infants on this basis (Almond et al., 2010). In order to utilise birth-weight alongside gestational age, its strong correlation with gestational age should be taken into account. The small for gestational age (SGA) classification is frequently employed. SGA is normally defined as being in the bottom 10% of birth weights for a gestational age week (Figure 1.1 in Chapter 1) (Carlo, 2011). Alternatively, a birth-weight z-score is used where birth-weights are normalised by gestational age week. I use the latter in this thesis.

The proportion of missing values for birth weight in the NNRD is practically zero for all infants born from 2010 onwards (and less than 2% for infants born prior to 2010) (Table 2.2). However, as Figure 2.3c shows, in the raw data there appear to be birth weights that are outliers for the gestational age week. This is likely to arise from human error in data entry. In particular, many of the apparently erroneous birth weight

entries are recorded as about 10% of the mean value for the respective gestational age week suggesting that the final digit of birth weight has been missed off. However, these infants are a random subset of all the infants, thus these data are likely to be MCAR.

Sex

The sex ratio of live births in the United Kingdom is 105.1 male births to 100 female births (Office of National Statistics, 2012). This compares to a sex ratio of neonatal unit admissions of 126.7 males to 100 females. This reflects the widely observed fact that male infants are more likely to experience an adverse clinical outcome than their female counterparts (Stevenson, 2000). In addition, 0.04% of all admissions were classified as indeterminate sex.[6]

Multiple Birth

Multiple births are associated with lower birth-weight and gestational age when compared to singleton births. As such, they are over-represented among admissions to neonatal units. In 2012, there were 15.9 maternities with multiple births for every 1,000 women giving birth in England and Wales (Office of National Statistics, 2012),[7] whereas 13.1% of all neonatal unit admissions were of infants from multiple births. The average gestational age of an admitted infant from a multiple birth is 33.3 weeks versus 36.9 weeks for a singleton infant.[8]

Antenatal Steroid Administration

Administration of antenatal steroids has become routine practise in Western countries for women with anticipated preterm labour. The use of antenatal steroids has been shown to reduce the risk of mortality and several other morbidities associated with preterm birth by promoting fetal lung development (Roberts and Dalziel, 2006). The NNRD data show that 22.9% of admissions received a full or partial course of antenatal

[6] Indeterminate sex is recorded for individuals that are not classifiable into either male or female sex on the basis of observed characteristics of the infant such as the gonads or genitals.

[7] This includes live births and still births.

[8] A t-test of the difference between these values yielded a p-value of <0.001.

steroids; however, a value for antenatal steroid administration was not recorded for 13.3% of infants (Table 2.2). The mean gestational age for infants for whom antenatal steroid administration was not recorded was 36.2 weeks, with 75% of infants with missing data being born at over 34 weeks gestation. Given that antenatal steroids are administered to women at risk of preterm labour and that the majority of missing data are in term admissions, this suggests that missing data are more likely to be among infants whose mothers did not receive a course of antenatal steroids. This may mean these data are MAR or MNAR.

Clinical Outcomes

The most widely used clinical outcome in studies of newborn health is mortality (see Lasswell et al. (2010), for example). Mortality, when used in studies of outcomes in neonatal units, can be defined in one of three ways: early neonatal mortality—death within 7 days post birth; neonatal mortality—death within 28 days post-birth; and, any in-hospital mortality—death prior to discharge home from neonatal care. In-hospital mortality does not capture deaths post-discharge deaths, however, infants are not discharged from hospital if there is a risk of mortality, moreover deaths post-discharge are unlikely to be amenable to the quality of neonatal healthcare and are therefore arguably not of interest to these analyses. The mean in-hospital mortality rate among all admitted infants between 2006 and 2013 was 1.8%, although this has varied between 1.5% and 2.0% over the years of the study data (Table 2.1). The mean neonatal mortality rate was 1.3% and the mean early neonatal mortality rate was 0.9%.

Other clinical outcomes are also considered in this thesis. Bronchopulmonary dysplasia (BPD) is a chronic lung condition most common among very preterm infants. BPD is classified into three forms on the basis of the requirement for supplemental oxygen. Mild BPD is defined as the requirement for supplemental oxygen for at least 28 days post birth, but not at 36^{+0} weeks postmenstrual age; moderate and severe BPD are both defined as the requirement for supplemental oxygen for at least 28 days post birth and at 36^{+0} weeks postmenstrual age; moderate and severe BPD are then further delineated by the percentage of oxygen provided by ventilation (Ehrenkranz et al.,

2005). Since the oxygen percentage is not available in the NNRD, where used in this thesis, 'BPD' is defined as infants with either moderate or severe BPD. The proportion of infants meeting this definition is 9.2% in the NNRD.

Surgery for necrotising enterocolitis (NEC) is also considered in this study. NEC is a disease of prematurity and is characterised by necrosis of the bowel (Lin and Stoll, 2006). Among very preterm infants, 1.4% received surgery for NEC. Infants born at term do not experience NEC.

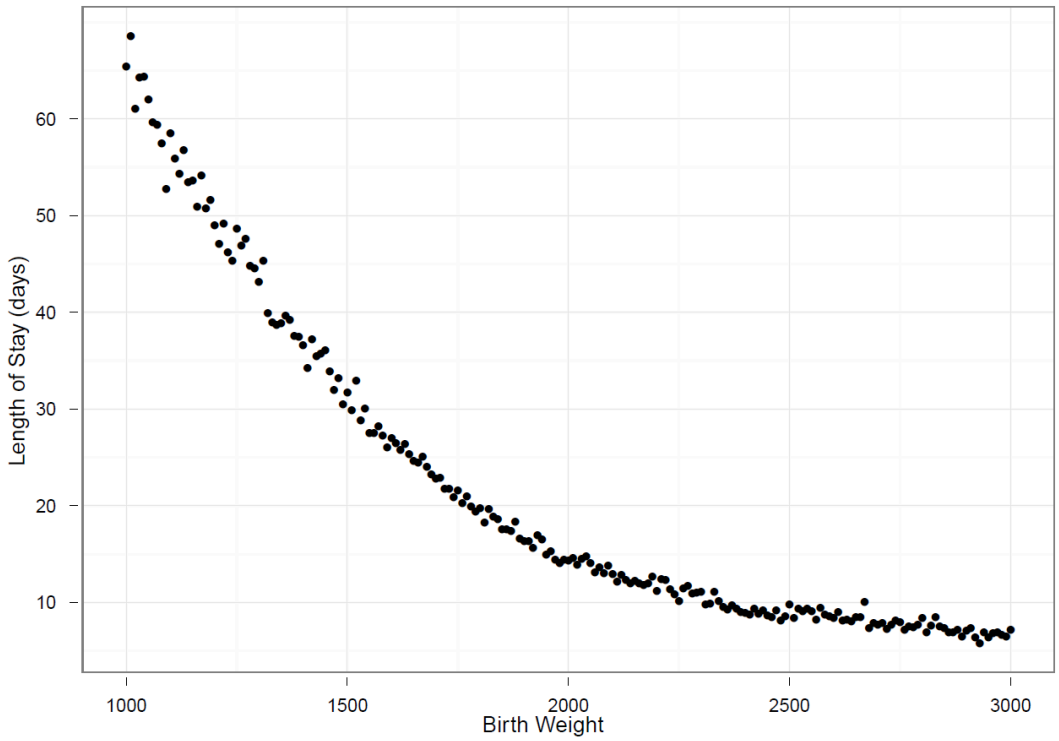
Retinopathy of prematurity (ROP) is a disease of the eye associated with prematurity and is generally seen among infants who received supplementary oxygen therapy (Stenson et al., 2013). Overall, 1.6% of very preterm infants in the NNRD received treatment for ROP.

Length of stay is an important outcome of interest because it both represents the clinical requirements of the individual infant and provides a marker of economic requirements of the infant. For example, Figure 2.4 shows how length of stay is strongly correlated with birth weight, and Table 2.3 shows its relationship with gestational age. These relationships are clearly non-linear suggesting the economic requirements of very low birth weight infants are disproportionately greater than their normal birth weight counterparts. The issue of healthcare requirements and the returns to healthcare expenditure is studied in Chapter 5.

2.2.4 Maternal Residence

The location of the mother's residence is a key variable in many of the analyses in this thesis. It allows for a linkage between infant characteristics and outcomes and local area socio-economic characteristics (see Section 2.3) and is also highly important in the empirical methodology where the nearest neonatal unit to the maternal residence is required (see Chapters 4, 5, and 6).

Fig. 2.4 Average length of stay by birth weight



Average length of stay is calculated for 10g bins of birth weight.

2.2.5 Maternal Variables

A number of maternal variables are available in the NNRD data including smoking, alcohol and drug use during pregnancy and maternal occupation. However, these are generally of poor quality and are missing in the majority of cases. Maternal age is also available and is of reasonable quality. These maternal variables are not generally used in models of neonatal clinical outcomes (Medlock et al., 2011), moreover exploratory analyses did not find evidence of an effect of these variables on infant clinical outcomes after conditioning on the previous clinical variables.

2.3 Population data

The majority of infants in the NNRD have the location of the maternal residence recorded (see Section 2.2.4). Among other things, this enables the linkage of an infant to geographical areas and hence socio-economic data related to those areas. A wide

Table 2.4 Super Output Areas in England

	N	Min pop.	Max. pop	Min households	Max. households
OA	171,371	100	625	40	250
LSOA	32,844	1,000	3,000	400	1,200
MSOA	6,791	5,000	15,000	2,000	6,000

[1] Source: ONS
[2] OA: Output area; LSOA: Lower layer super output area; MSOA: Middle layer super output area

range of data on local social and economic variables are available for England and Wales from the Office for National Statistics (ONS). These data are used widely in this thesis. Socio-economic status has been previously shown to be associated, after controlling for relevant and observable clinical factors, with the requirements for neonatal healthcare (Smith et al., 2007, 2009). As such, local area deprivation is included in many of the structural models estimated in this work, the measure of deprivation is described in Section 2.3.2. Furthermore, in Chapter 7, I estimate the effect of local area economic conditions on infant health at birth.

2.3.1 Geographical Areas

To facilitate the reporting of census data, the United Kingdom is divided into Output Areas (OAs), which are designed to have similar population sizes and be as socially homogeneous as possible. OAs are further grouped into Lower Layer Super Output Areas (LSOAs) and Middle Layer Super Output Areas (MSOAs). Numbers and definitions of different OAs in England are reported in Table 2.4.

2.3.2 Socio-economic Deprivation

The association between poor health and low socio-economic status has been widely documented (Currie, 2009; Mackenbach et al., 2008). There are many factors that may mediate this relationship, such as education, local environmental conditions, and access to healthcare. These mechanisms are described in Chapter 1 (see, in particular, Figure 1.2.) These data are not observed for individual parents in the NNRD. However, based on the maternal residence, local area socio-economic deprivation can be assigned

to each infant.

A widely used measure of local area socio-economic deprivation is the Index of Multiple Deprivation (IMD) created for the Department of Communities and Local Government. The IMD is a relative measure of deprivation based on 38 indicators grouped into seven domains: income, employment, health, education, crime, access to services and living environment (Noble et al., 2007). The IMD is reported for LSOAs. As the index is ordinal, local areas are typically divided up into quintiles or deciles for use in statistical models. In this study areas are divided into socio-economic quintiles where the IMD is used.

Within in the NNRD, infants from the most deprived LSOAs are over-represented compared to the general population, which is consistent with the previously reported association between socio-economic status and poor health. Specifically, 28.1% of the admissions in the NNRD are from the most deprived quintile of LSOAs compared to 14.3% from the least deprived quintile. Table 2.5 shows descriptive statistics for the infants in the NNRD by quintile of socio-economic deprivation. Infants admitted whose mothers lived in the most deprived areas are, on average, earlier term (36.3 weeks for infants born to mothers living in the most deprived areas compared to 36.6 weeks for the least deprived areas) and of lower birth weight (2,669.9g for the most deprived quintile versus 2,831.4g for the least deprived quintile) than their counterparts born to mothers from lower deprivation areas. This translates into different rates of mortality and BPD between areas by socio-economic deprivation. For example, the mortality rate among infants born to mothers residing in the most socio-economically deprived areas is 2.1% compared to 1.4% for infants born to mothers residing in the least deprived areas. The corresponding BPD rates are 9.6% and 7.8%. Nevertheless, the work by Smith et al. (2009) suggests that "The burden of mortality and morbidity is greater among babies born to women from deprived areas because of increased rates of very preterm birth. After very preterm birth, however, survival rates and neonatal care provision is similar for infants from all areas." (see abstract in Smith et al. (2009)).

Table 2.5 Descriptive statistics of infants by quintile of deprivation

Variable	Deprivation quintile					P-value[a]
	1 (Most deprived)	2	3	4	5 (Least deprived)	
Gestational age (wks)	36.3(3.9)	36.5(3.8)	36.6(3.8)	36.6(3.7)	36.6(3.7)	<0.001
Birth weight (g)	2,669.9(917.7)	2,741.7(922.5)	2,791.8(918.9)	2,817.7(918.2)	2,831.4(907.2)	<0.001
Antenatal steroids (%)	25.0	24.0	23.3	23.3	23.4	<0.001
Male (%)	55.3	55.5	56.1	56.6	56.8	<0.001
Multiple birth (%)	10.8	12.4	13.9	15.2	16.9	<0.001
Mothers age (yrs)	28.4(6.3)	29.5(6.2)	30.4(6.1)	31.3(6.0)	32.3(5.8)	<0.001
Mortality (%)	2.1	1.9	1.6	1.5	1.4	<0.001
BPD (%)	9.6	9.0	8.5	8.3	7.8	<0.001

[1] Values are mean (sd) unless otherwise stated.
[2] Socio-economic deprivation quintiles are based upon the Index of Multiple Deprivation rank of the LSOA.
[a] P-values from ANOVA test for continuous variables, and chi-squared test for categorical variables.

2.3.3 Unemployment Rate

The unemployment rate is a key variable for the analysis in Chapter 7 and is discussed in greater detail there. The unemployment rate may be defined in a number of ways. The primary definition used in this thesis is the claimant count rate, defined as the percentage of the working age population claiming Jobseeker's Allowance (JSA).[9][10] JSA is a state administered payment to individuals meeting the following criteria: aged over 18 but below the state pension age, not in full time education, able and available to work, and actively seeking work. The individual must work on average less than 16 hours per week and have less than £16,000 savings. The unemployment rate is available at the LSOA and MSOA levels and for all years between 2006 and 2013.

2.3.4 Area Level Birth Rate and Birth Weight

The NNRD contains information only on those infants admitted to neonatal specialist healthcare and so it cannot be used alone to make inferences about population level infant health. The total number of live births per MSOA are available from the ONS. This can be used to derive an admissions rate per MSOA. These data are shown in Table 2.6. The admission rate increased from 7.6% of all live births in 2007 to 9.0% in 2011. This is reflected by the increased number of admissions per unit observed in Table 2.1.

The admission rate can be used to infer the health of the birth cohort in any given year, as I do in Chapter 7; nonetheless, the most common measure of health at birth is birth weight. Data were obtained from the ONS on the birth weights of infants by MSOA, categorised into 500g birth weight categories for the years 2006-11.[11] Data relating to smaller areas, such as LSOAs, were not available since these may be identifiable given the small number of very low birth weight infants. The total number of births as well as the birth rate has increased over the period of the NNRD as Tables

[9]The working age population is typically defined as those aged between 16-65, however, due to the demographic data available from the ONS, the working age population used here is those aged 15-64.

[10]The JSA rate was £71.70 per week for an individual over 25 in 2013, and £56.80 for an individual aged 16 to 24.

[11]Data more recent than this were not available when the data were obtained in January 2013.

Table 2.6 Summary statistics for numbers of births by birth weight and proprtion of admissions

Variable	**2007**	**2008**	**2009**	**2010**	**2011**
Births (N)	655,357	672,809	668,678	687,006	688,119
<1500g	7,482(1.1)	8,044(1.2)	7,936(1.2)	8,171(1.2)	7,982(1.2)
1500-2000g	9,337(1.4)	9,507(1.4)	9,546(1.4)	9,461(1.4)	9,542(1.4)
2000-2500g	29,954(4.6)	30,282(4.5)	30,294(4.5)	30,067(4.4)	30,816(4.5)
>2500g	601,324(91.8)	619,630(92.1)	618,334(92.5)	632,538(92.1)	633,609(92.1)
Admissions (%)	7.6	7.9	8.3	8.4	9.0

[1] Source: ONS and NNRD, 2006-13
[2] Birth weight rows are the number of births the that category (%).
[3] The admissions rate is the percentage of live births admitted onto a neonatal unit

Table 2.7 Fertility and birth statistics for live births in England and Wales

Year	Total Fertility Rate[a]	General Fertility Rate: all live births per 1,000 women aged 15-44	Crude Birth Rate: all births per 1,000 population of all ages
2012	1.94	64.8	12.9
2011	1.93	64.0	12.9
2010	1.94	64.0	13.0
2009	1.90	62.5	12.8
2008	1.92	62.7	12.9
2007	1.88	61.2	12.7
2006	1.83	59.6	12.4

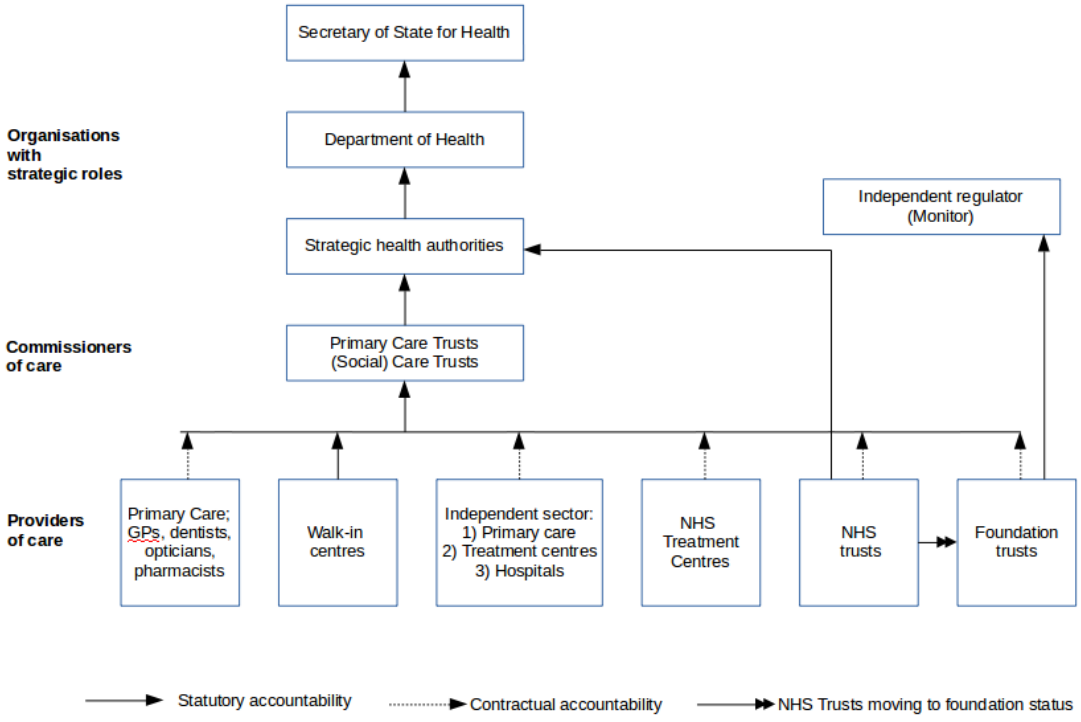
[1] Source: ONS
[a] The Total Fertility Rate is the average number of live children that a group of women would bear if they experienced the age-specific fertility rates of the calendar year in question throughout their childbearing lifespan.

2.6 and 2.7 show respectively. However, the proportion of these births that are VLBW has not changed, remaining at 1.2% between 2007-11. Therefore, the increase in admissions has been of higher birth weight and term infants; again this is shown in Table 2.1.

2.4 National Health Service data

The structure of the NHS was briefly outlined in Chapter 1. To provide more detail, Figure 2.5 shows the structure of the NHS in England between April 2003 and April 2013. Of key interest, are the commissioners and providers of care, in particular Pri-

Fig. 2.5 The structure of the NHS in England, 2003-13.



Adapted from Talbot-Smith and Pollock (2006)

mary Care Trusts (PCTs) and NHS Trusts and Foundation Trusts. Between October 2006 and April 2013 there were 152 PCTs covering the population of England (down from 303 between April 2002 and October 2006—PCTs were replaced by Clinical Commissioning Groups (CCGs) in April 2013). PCTs were responsible for identifying health needs of their geographically defined population and securing appropriate healthcare provision in primary, secondary, tertiary (or specialist), and community healthcare sectors. PCTs commissioned secondary and tertiary healthcare services from NHS Trusts (a generic name given to variety of hospital trusts) and Foundation Trusts (FT). FTs are notionally independent organisations not under the direction of the Secretary of State for Health unlike other NHS Trusts. They contracted with PCTs to provide healthcare with legally binding contracts, functioning as corporate institutions, unlike other NHS Trusts which use NHS contracts that are not legally binding but are enforceable by the Secretary of State. NHS FTs were able to apply for funds from their local Strategic Health Authority (SHA) to support capital investment. In April 2013 there were 147 FTs.

Two sources of data are obtained from the NHS: the Programme Budgeting Data which provides estimates of PCT level expenditure across 23 programmes of healthcare, and the Reference Costs which provide estimates of the unit costs of providing healthcare across all Healthcare Resource Groups (HRGs). HRGs are classifications used in the NHS used to categorise patients who consume similar levels of resources, for example, neonatal intensive care days. These data are primarily used in Chapter 5, where the returns to healthcare expenditure in neonatal specialist care are estimated. Previous studies have utilised the Programme Budgeting Data to do exactly this (Claxton et al., 2013; Martin et al., 2008); however, the reference cost data has not been previously used for this purpose.

2.4.1 Programme Budgeting Data

As previously outlined, until 2013 PCTs were responsible for the commissioning of primary, secondary, and community healthcare (Talbot-Smith and Pollock, 2006). The

Programme Budgeting Data contain annual, PCT estimates of the total expenditure across 23 programmes of healthcare. The programmes of care represent the various specialities of healthcare provided by the NHS in England and are defined by International Classifications of Diseases, Version 10 (ICD-10) codes. Every patient is assigned at least one of these codes and on this basis, care provided to a patient is assigned into a particular programme of care,[12] one of which is neonatal healthcare. The determination of costs by providers occurs in two steps: firstly, the level of activity in all HRGs is determined, and secondly, the provider estimates unit costs for each HRG and then, using the activity, calculates total costs. The PCTs combine information from all of its providers to determine its expenditure (Department of Health, 2011; NHS England, 2013).

Table 2.8 shows summary statistics from the programme budgeting data. The total expenditure across all PCTs on neonatal care was approximately £1 billion annually, which constituted approximately 1% of PCT expenditure. This translated into a cost per head of population of £18.50 in 2011. The average cost per neonatal care day in 2011 was £1,435.9.

2.4.2 Reference Costs

The reference costs are provider averages of the estimated costs of providing care within each HRG. The Department of Health use these reference costs to set prices for NHS funded services in England. The raw data at the provider level, used to calculate the average costs, are publicly available. Each provider estimates their cost for each HRG by taking into account fixed costs (e.g. depreciation), semi-fixed costs (e.g. nursing staff), and variable costs (e.g. drugs and consumables) (Monitor, 2014). Thus, these costs represent different levels of factor inputs to healthcare production and can

[12]The programmes of care are: infectious diseases; cancers and tumours; disorders of the blood; endocrine, nutritional, and metabolic problems; mental health disorders; problems of learning disability; neurological; problems of vision; problems of hearing; problems of circulation; problems of the respiratory system; dental problems; problems of the gastro-intestinal system; problems of the skin; problems of the musculo-skeletal system; problems due to trauma and injuries; problems of genito-urinary system; maternity and reproductive health; conditions of neonates; adverse effects and poisoning; healthy individuals; social care needs; other. Many of these categories are further divided into subcategories for which expenditure data is available.

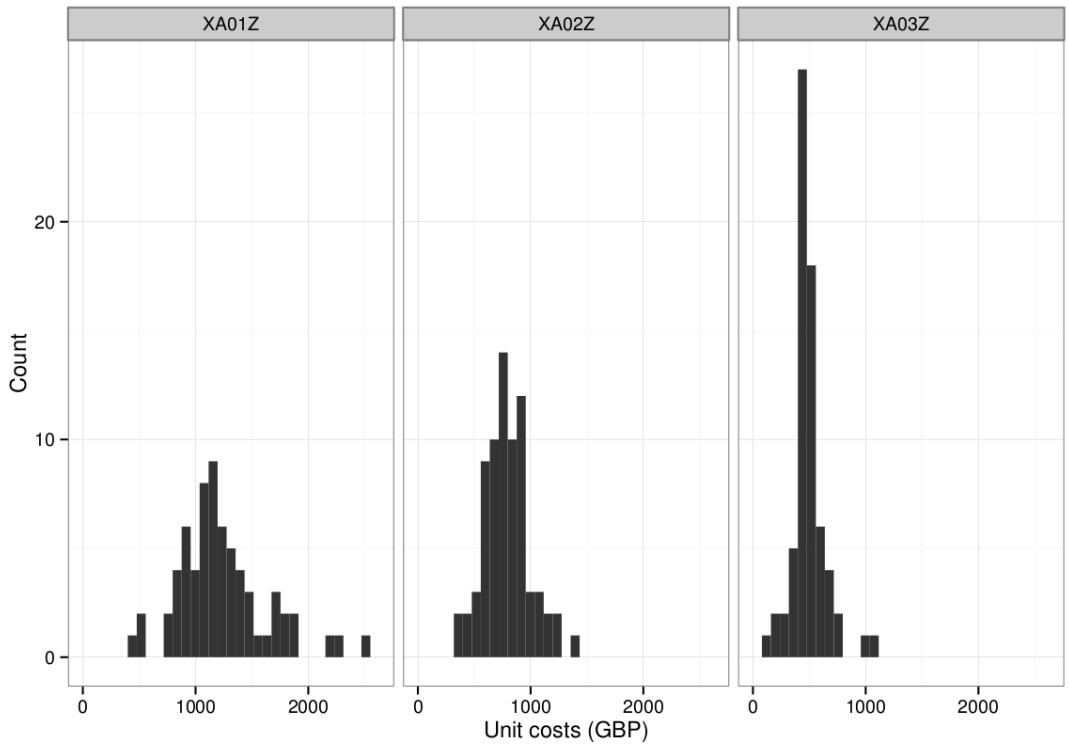
Table 2.8 Summary statistics for programme budegting data

Variable	**2008/9**	**2009/10**	**2010/11**	**2011/12**	**2012/13**
Total expenditure (£million)	882.346	1,012.833	834.606	936.163	981.354
% total PCT expenditure	1.12	1.16	0.91	1.01	1.04
Expenditure per capita (2011£)	18.9	20.6	16.5	16.8	18.5
	(6.1)	(8.1)	(6.3)	(6.9)	(3.6)
Expenditure/admission (2011£)[a]	18,263.3	19,783.3	16,450.9	15,459.1	15,658.1
	(7,110.9)	(9,233.3)	(6,859.0)	(7,309.0)	(6,964.7)
Expenditure/careday (2011£)[a]	1,763.7	1,842.5	1,480.9	1,435.9	1,542.1
	(621.5)	(552.3)	(565.4)	(558.5)	(515.0)

[1] Source: ONS

[a] Total number of admission and care days were not available for all PCTs to calculate expenditure per admission and care day. PCTs were excluded from these calculations on the basis on Chauvenet's criterion. As has been discussed in this chapter (Section 2.2.1) not all neonatal units contribute data to the NNRD which means that not all admissions and care days provided to infants originating from certain PCTs are observed. Thus the denominator in the expenditure per admission of per care day calculations are complete. In order to derive a figure for average expenditure per admission or care day, it is necessary to exclude those PCTs which are not completely observed, assuming that expenditure per admission and per care day are normally distributed, Chauvenet's criterion is an algorithm that succesively eliminates outlying values until the remaining value are normally distributed.

Fig. 2.6 Distribution of estimated unit costs of neonatal healthcare by care level



Source: NHS Reference Costs. XA01Z=Intensive care, XA02Z=High dependency care, XA03Z=Special care.

be used to estimate the effects of increasing these inputs, in monetary terms. However, the number of providers contributing data in each year varies with 125 for the financial year 2008/9, 128 in 2009/10, 80 in 2010/11 and 2011/12, and 73 in 2012/13.

Figure 2.6 shows distributions in the estimated unit costs for the three most frequently used HRGs within neonatal care from the providers contributing in 2012: intensive care cot days, high dependency care cot days, and special care cot days.[13] The mean (sd) unit cost for an intensive care day was £1,232.66 (395.14). The respective figures for high dependency care day and special care day were £793.53 (205.76) and £496.09 (148.60), respectively. Note that these figures are all below the average cost per care day shown in Table 2.8 which was calculated using the programme budgeting data. This is due to the fact that the health related groups for which the reference costs are estimated are homogeneous populations and so do not take into account extraordi-

[13]The definitions of each of the types of care days are provided in British Association of Perinatal Medicine (2011)

nary or uncommon and expensive procedures such as surgery and neonatal transfers. Thus, the programme budgeting data may reflect population health within the PCT, whereas the reference costs represent differences in provider inputs to neonatal healthcare and their respective prices. This difference is further discussed in Chapter 5.

Alongside the provider level estimates of unit costs, the data contain estimated market forces factors (MFF) for each provider. The MFF is an index designed to capture unavoidable differences between healthcare providers in factor inputs to healthcare production including labour, capital inputs such as land and buildings, and a London weighting (Monitor, 2013). The MFF varied between 0.920 and 1.298, where a value of 1.100 for an area means that the unavoidable costs are 10% higher in that area compared to an area with an MFF of 1.000. This index is used in the models of healthcare expenditure in Chapter 5.

2.5 Unit Profile Survey 2011

The Unit Profile Survey 2011 (UPS) was a survey of English neonatal units conducted in November 2011. Of 171 units surveyed, 159 (93.0%) responded. The survey aimed to collect data on labour, including nursing and physician staffing both in post and establishment, and capital, such as cots and surgical facilities. It was also intended to capture the exact labour inputs on a specific day of the year to provide a cross-section of inputs to neonatal care to match to data on infants born in and admitted to neonatal care on that day. However, not all questions were completed to a high quality. The UPS was a follow up to two previous surveys conducted by the Medical Research Council EPICure studies in 1997 and 2006 (Hamilton et al., 2007; Tucker, 2002). Further details are given in Chapter 5 where the number of whole time equivalent (WTE) staff along with the number of cots per care day are compared to neonatal unit expenditure per care day.

Chapter 3

Literature Review

The goal of this chapter is to survey the literature relevant to the topics of this thesis with the view to providing a clear overview of the fields of study. This chapter has two main foci: research outcomes and research methods. In the former case, integration of the literature from the standpoint of outcomes provides insights into the current state of knowledge in the fields in order to delimit the research questions posed in this thesis and identify where new contributions can be made. A review of research methods is crucial for a number of reasons, including identifying key variables and, more importantly, the methods used in prior analyses. Previous methodological approaches to the research questions in the relevant literature can be appraised critically in order to guide the methods used in this thesis. There is likely to be some disparity in the approaches used since studies feature from a number of disciplines including clinical science, economics, and statistics.

It is also important to understand the processes underlying any observed results. This includes the practices of staff within labour units, the institutional organisation of the healthcare systems, as well as the social process producing aggregate outcomes. This knowledge is important to guide the methodology used and the interpretation of results.

This thesis broadly aims to examine the effects of the organisation of neonatal healthcare has on the clinical and economic outcomes of newborn infants treated within those units. An increasing body of research is devoted to understanding the link be-

tween unit characteristics and outcomes within neonatal care. In England and Wales, neonatal mortality comprised 34% of the overall infant mortality (under one year) rate in 2008, indicating that this group is a particularly vulnerable one (Office of National Statistics, 2011). Moreover, there has been an increase in the rate of preterm birth in recent years (Goldenberg et al. (2008); Zeitlin and Ancel (2011), see also Chapter 2).

As this chapter shows, unit volume, measured either in terms of the number of patients admitted or caredays provided per annum, is the most frequently examined unit characteristic. However, it is essential not to overlook other important unit variables. Variables such as staff levels, occupancy, and resourcing may be crucial to understanding how unit level differences affect individual outcomes. Indeed, guidelines for neonatal units in England (e.g. Department of Health (2003); National Audit Office (2007)) cite one study (Tucker, 2002) that indicates that higher levels of occupancy at admission lead to worse clinical outcomes on neonatal units.

Non-healthcare related determinants of infant health, including social and economic factors, are also of high interest since these can affect the size and composition of the birth cohort. Any changes in the health of the birth cohort has large effects on the effectiveness of, and costs associated with, neonatal healthcare. Thus, policy makers aiming to improve newborn health may be able to target interventions further upstream, at the social and economic determinants of poor infant health, to a potentially greater effect than policies aimed at neonatal unit characteristics. These socio-economic aspects are identified in this chapter where they appear in the included studies.

The following section outlines the methods for this review, including search terms and databases. Section 3.2.1 explores the results of the research identified in the review, while Section 3.2.2 examines the research methods used in previous studies.

3.1 Methods

The following relevant databases were searched: PUBMED, Google Scholar, EconLit, and IDEAS. Combinations of terms relating to neonatal care (e.g. newborn care, neonatal, neonatal unit, intensive care unit, special care unit etc.), unit or network

characteristics (e.g. admissions, volume, staffing etc.), and outcomes (e.g. mortality, morbidity, costs etc.) were searched. Table 3.1 shows the search terms. The limits to the searches were: a) 1990 onwards, b) English language, c) study used a statistical or econometric methodology,[1] d) reported an effect of a unit characteristic on an individual outcome, e) reported a neonatal unit characteristic,[2] and f) the full paper was available. Bibliographies of selected studies were also searched. Other papers were included where appropriate, such as those identified in other systematic reviews (see Lasswell et al. (2010); Medlock et al. (2011); Sherenian et al. (2013)).

3.2 Results from the Literature Review

Table 3.1 shows the search terms and number of citations found and selected. Appendix A provides a table summarising all papers that are included in the review. Of the 35 studies that met the inclusion criteria, 28 reported on clinical outcomes,[3] six reported on economic outcomes,[4] and one reported on both [5]. Mortality was the most frequently studied clinical outcome with 24 studies reporting a mortality outcome, of which 13 examined neonatal mortality (death within 28 days)[6] and four studied early neonatal mortality (death within seven days).[7] The other studies were either not explicit about the timing of mortality or used one year outcomes (that is death within one year for infants who remained hospitalised over this period). The most frequently reported

[1]This limit was used to select only quantitative studies; while important, qualitative studies were excluded as they were not relevant to the research conducted in the thesis, and reviews did not contain original research.

[2]Many studies were found when searching that related to the volume of obstetric care or the number of deliveries in a hospital.

[3]Abdel-Latif et al. (2006); Baker and Phibbs (2002); Bartels et al. (2006); Bell et al. (2010); Bode et al. (2001); Callaghan et al. (2003); Chung et al. (2011, 2010); Cifuentes et al. (2002); Cimiotti et al. (2006a); Field and Draper (1999); Filho et al. (2011); Goodman et al. (2002); Grandi et al. (2010); Hamilton et al. (2007); Horbar et al. (1997); Lake et al. (2012); Lorch et al. (2012); Phibbs et al. (2007, 1996); Pollack and Koch (2003); Profit et al. (2006); Rogowski et al. (2004); Shim et al. (2013); Straney et al. (2012); Synnes et al. (2006); Tucker (2002); Wall et al. (2004)

[4]Fordham et al. (1992); Hollingsworth and Parkin (2001); Leleu et al. (2012); O'Neill and Largey (1997); O'Neill et al. (2000); Roblin et al. (2000)

[5]Almond et al. (2010)

[6]Baker and Phibbs (2002); Bartels et al. (2006); Bell et al. (2010); Bode et al. (2001); Chung et al. (2011, 2010); Cifuentes et al. (2002); Goodman et al. (2002); Horbar et al. (1997); Lake et al. (2012); Phibbs et al. (2007, 1996); Pollack and Koch (2003)

[7]Abdel-Latif et al. (2006); Bell et al. (2010); Lake et al. (2012); Straney et al. (2012)

Table 3.1 Results from literature review searches

Search Term	Citations found	Citations selected (incl. duplicates)	Citations selected (excl. duplicates)
PUBMED			
{neonatal unit}[tiab] AND admissions[tiab] AND mortality[tiab]	24	1	0
{neonatal unit}[tiab] AND volume[tiab] AND mortality[tiab]	6	1	1
NICU[tiab] AND volume[tiab] AND mortality[tiab]	27	7	6
NICU[tiab] AND (staff[tiab] OR staffing[tiab]) AND mortality[tiab]	28	2	1
neonatal[tiab] AND (occupancy[tiab]) AND mortality[tiab]	11	1	0
neonatal[tiab] AND {patient volume}[tiab] AND outcome[tiab]	12	7	2
neonatal[tiab] AND {staffing}[tiab] AND outcome[tiab]	29	6	3
hospital[ti] AND very low birth weight[tiab] AND intensive care[tiab]	41	4	4
EconLit			
Neonatal (abstract)	86	7	3
IDEAS			
Neonatal outcome	60	2	1
Google Scholar[a]			
neonatal unit outcome volume	20	2	0
neonatal unit outcome staff	20	3	4
neonatal unit variables patient outcome mortality	20	4	2
neonatal unit (regionalization OR deregionalization)	20	5	3
Other	-	-	12
Total	398	48	35

[a] Only the first twenty results from Google Scholar are used given the large number of results

Table 3.2 The number of studies reporting various unit characteristics and outcomes

		Outcome				
		Mortality	Morbidity	Length of Stay	Transfer Status	Other economic outcome
Variable	Volume	11	2	1	1	4
	Level of care	4	1	-	1	1
	Staffing provision	9	6	-	-	-
	Other	1	1	2	-	-

A study may appear more than once if multiple characteristics or outputs were considered. A total of 28 studies were included.

morbidity outcome was intraventricular haemorrhage (IVH), used as an outcome in six studies.[8] Only one study used transfer status as an outcome (Wall et al., 2004). For the economic outcomes, three studies estimated a production function (Fordham et al., 1992; O'Neill and Largey, 1997; O'Neill et al., 2000), one used length of stay to determine resource utilisation (Roblin et al., 2000), one study used number of care days to determine returns to scale (Leleu et al., 2012), and the one examined technical efficiency of neonatal units in England (Hollingsworth and Parkin, 2001).

The majority of studies identified were published in medical journals. These studies generally report odds ratios from logistic regression as is the standard in the medical literature. Furthermore these studies do not generally focus on causal inference as is the practice in economic literature. Econometric and statistical issues are explored in Section 3.2.2.

[8]Bell et al. (2010); Grandi et al. (2010); Lake et al. (2012); Pollack and Koch (2003); Profit et al. (2007); Synnes et al. (2006)

Location of Studies

The majority of the studies (20) were from the US [9], of which eight used only Californian data,[10] seven were from the United Kingdom,[11] one was from Germany (Bartels et al., 2006), one from Canada (Synnes et al., 2006), two from Australia (Abdel-Latif et al., 2006; Callaghan et al., 2003), one from Korea (Shim et al., 2013), and two from South American nations (Filho et al., 2011; Grandi et al., 2010).

There is an important question as to whether results from outside of the UK are generalisable to the UK. The structure of neonatal care in the United States is somewhat different to that for the UK with the most well represented area being California with eight studies coming from there.[12]

The categorisation of care level in the United Kingdom is used to inform the pathways of care for babies admitted to neonatal care. As has been described in the previous chapters, neonatal units are arranged in the United Kingdom into managed clinical networks (now called operational delivery networks) which generally consist of one or two intensive care (level 3 units) in the centre of a network comprising a number of other lower level units. Within these networks babies are transferred to a unit that can provide an appropriate level of care if they are not already in one. The ethos is that it is the network that provides the care and not the individual unit. This does not mean that there are no transfers in neonatal care in the United States. However there is a paucity of data that describes transfers, and, one might assume, the cooperation and coordination required in the health maintenance organization (HMO) system and distances in the US may preclude the formation of managed clinical networks there.

If one considers babies by their place of birth, the casemix of babies found in low

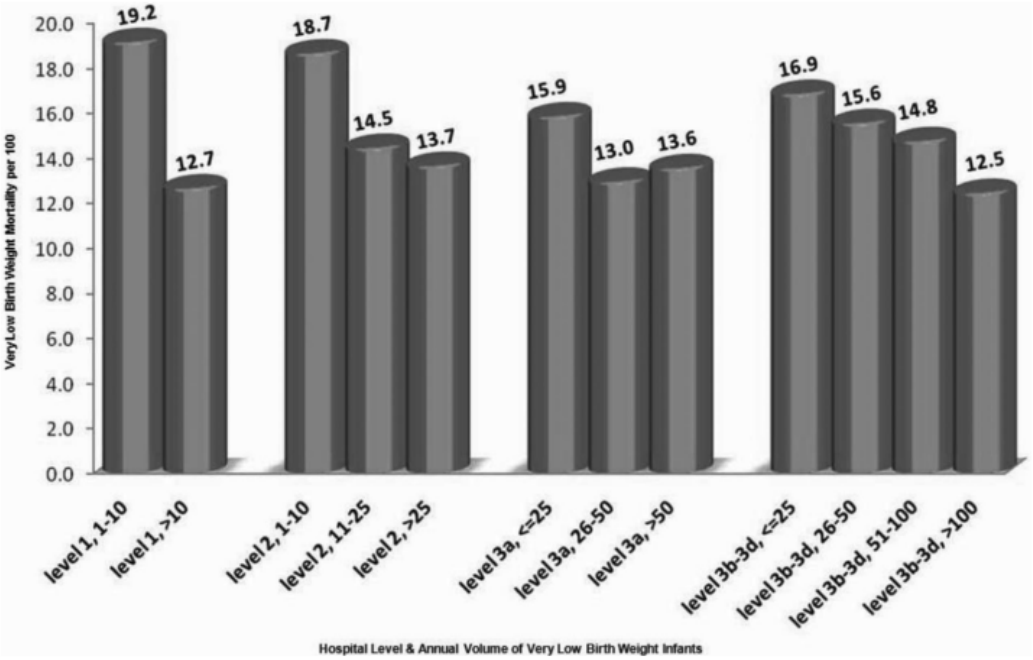
[9] Almond et al. (2010); Baker and Phibbs (2002); Bell et al. (2010); Bode et al. (2001); Chung et al. (2011, 2010); Cifuentes et al. (2002); Cimiotti et al. (2006a); Goodman et al. (2002); Horbar et al. (1997); Lake et al. (2012); Leleu et al. (2012); Lorch et al. (2012); Phibbs et al. (2007); Pollack and Koch (2003); Profit et al. (2010); Roblin et al. (2000); Rogowski et al. (2004); Straney et al. (2012); Wall et al. (2004)

[10] Baker and Phibbs (2002); Chung et al. (2011, 2010); Cifuentes et al. (2002); Lorch et al. (2012); Phibbs et al. (2007); Profit et al. (2010); Roblin et al. (2000)

[11] Fordham et al. (1992); Gale et al. (2012b); Hamilton et al. (2007); Hollingsworth and Parkin (2001); O'Neill and Largey (1997); O'Neill et al. (2000); Tucker (2002)

[12] Baker and Phibbs (2002); Chung et al. (2011, 2010); Cifuentes et al. (2002); Lorch et al. (2012); Phibbs et al. (2007); Profit et al. (2010); Roblin et al. (2000)

Fig. 3.1 Crude mortality rates by unit level and volume of VLBW admissions for units in California 1997-2002, from Chung et al., (2010)



care level hospitals in the UK is different to the US. For example, examining data presented in the US studies by Chung et al. (2011, 2010) and Phibbs et al. (2007), shows that the crude mortality (mortality unadjusted for patient characteristics) rates in low level hospitals are higher than those for higher level hospitals (see Figure 3.1). This would be an unexpected finding if high risk babies were transferred in utero to higher level units where they could receive more appropriate care. Samuelson et al. (2002) speculated that around 16-26% of neonatal deaths in Georgia, USA could be prevented if VLBW babies were delivered in level three centres. A recent study in the UK found that the formation of MCNs was associated with an increase in pre- and postnatal transfers of infants (Gale et al., 2012b). The UK benefits from a nationalised health service which has resisted regionalisation of neonatal care, opting instead for a networked approach as described in Chapters 1 and 2. However, it remains to be seen whether the same effect of unit volume is observed in the UK. Even between countries in Europe there are large differences in preterm survival (Saigal and Doyle, 2008a).

As this thesis will examine data from England, those studies that are from the United Kingdom are of particular interest. Three UK studies examined clinical out-

comes, none of which found a statistically significant relationship between the volume of a neonatal unit and the risk of mortality (Field and Draper, 1999; Hamilton et al., 2007; Tucker, 2002). The other four UK based studies looked at economic outcomes, three of which concluded that there were economies of scale present in English neonatal units although these were dependent on the proportion of intensive care provided (Fordham et al., 1992; O'Neill and Largey, 1997; O'Neill et al., 2000). The other calculated technical efficiency for English neonatal units(Hollingsworth and Parkin, 2001). Perhaps the most interesting results for the UK focussed on the relationship between staffing and neonatal outcomes. As has been previously mentioned, Hamilton et al. (2007) found that an increase in the specialist nurse to patient ratio was related to a reduction in mortality. Furthermore, the authors found that 57% of nursing shifts were understaffed (relative to BAPM guidelines). In a paper widely cited in UK neonatal guidelines, Tucker (2002) found that infants admitted at full capacity versus half capacity were about 50% more likely to die, although, as the authors admit, there was a lot of uncertainty around this estimate (see Figure 3.2). Even though there is some strong research from the UK, there is a paucity of research in this area. Moreover, with the exception of the study by Gale et al. (2012b) who did not examine clinical outcomes, none of the data used are from after the formation of MCNs.

3.2.1 Review of Findings of Included Studies

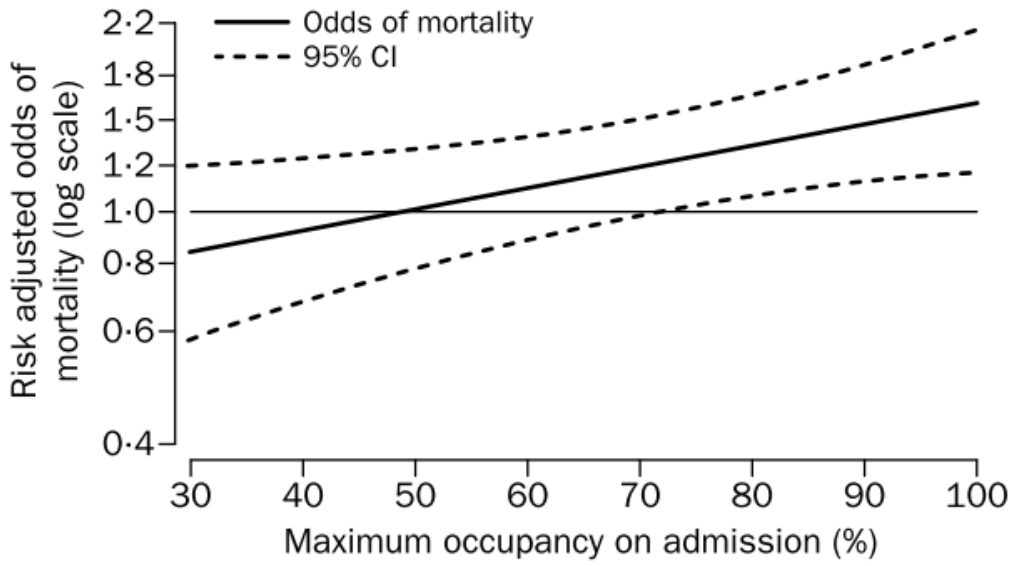
The methods and results of individual studies discussed in this section are provided in Appendix A.

Volume

Patient volume was studied as a unit characteristic in 20 studies.[13] It was measured in one of two ways, either by number of cots or by number of patients, and in all cases

[13] Bartels et al. (2006); Chung et al. (2010, 2009); Cifuentes et al. (2002); Field and Draper (1999); Fordham et al. (1992); Hamilton et al. (2007); Hollingsworth and Parkin (2001); Horbar et al. (1997); Leleu et al. (2012); O'Neill and Largey (1997); O'Neill et al. (2000); Phibbs et al. (2007); Roblin et al. (2000); Rogowski et al. (2004); Shim et al. (2013); Straney et al. (2012); Synnes et al. (2006); Tucker (2002); Wall et al. (2004)

Fig. 3.2 Estimated risk of mortality by maximum occupancy at admission with 95% confidence interval from Tucker (2002)



the volume was an annual measure. Typically, the patients counting towards measures of unit volume were from a subset of the patient population, such as, very low birth weight (<1,500g; VLBW) infants. For example, Bartels et al. (2006) examined the relationship of NICU and delivery unit volume in relation to mortality for infants born between 24 and 30 weeks gestation, but used the annual number of VLBW admissions to define neonatal intensive care unit (NICU) volume. Indeed, ten of the 20 studies examining volume defined volume as the annual number of VLBW admissions;[14] two others used the annual number of low birth weight (<2,500g; LBW) admissions as a measure of volume (Cifuentes et al., 2002; Hamilton et al., 2007).

A widely cited paper on this topic, Phibbs et al. (2007), found that level 3B, 3C, or 3D units (considered as one group) with an annual admission of greater than 100 VLBW babies in California, US, were associated with statistically significant reductions in mortality compared to all other units.[15] For example, the odds ratio for mor-

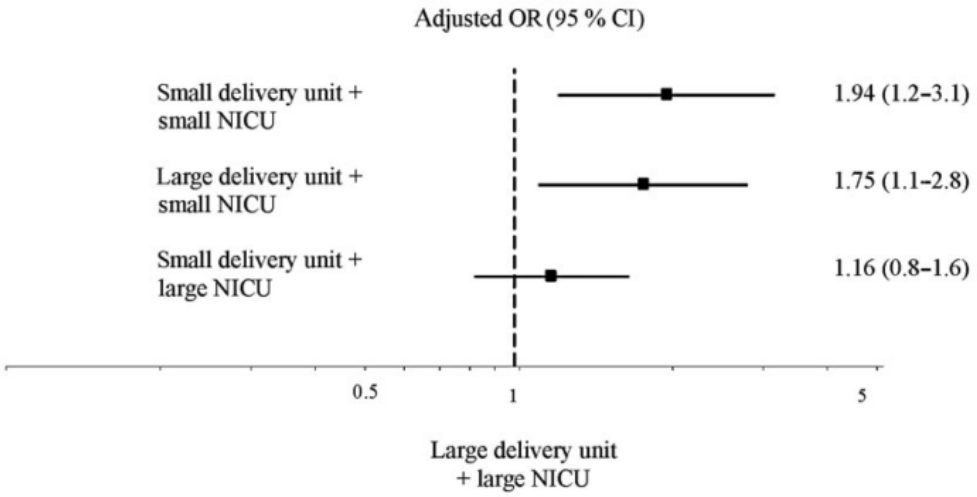
[14] Bartels et al. (2006); Chung et al. (2011, 2010); Horbar et al. (1997); Phibbs et al. (2007); Rogowski et al. (2004); Shim et al. (2013); Synnes et al. (2006); Tucker (2002); Wall et al. (2004)

[15] These levels correspond to the following definitions: (1) no NICU, (2) NICU that cares for mildly ill infants but without mechanical ventilation, (3a) NICU that provides mechanical ventilation with restrictions (e.g. only below 1000g), (3b) NICU with mechanical ventilation with no restrictions but no major surgery facilities, (3c) NICU with mechanical ventilation and major surgery but neither open-heart surgery nor extracorporeal membrane oxygenation (ECMO), (3d) NICU with major surgery facilities requiring ECMO.

tality in a level 3B, 3C, or 3D unit with annual VLBW admissions of 51-100 was 1.19 (95% C.I. 1.04-1.37, p=0.01) compared to similar units with greater than 100 admissions. Chung et al. (2010) found a similar effect size for the same region, the associated odds ratio for level 3B, 3C, or 3D units with annual VLBW admissions of 51-100 was 1.32 (95% C.I. 1.15-1.52) compared to units with greater than 100 annual VLBW admissions. These results are the population averaged effect as opposed to the individual effect since clustering—outcomes of infants treated in the same unit may be correlated—at the unit level was not accounted for. Both studies combined all the data and did not account for the possibility that observations from the same unit are not independent; this form of model is known as a pooled model. Chung et al. (2011) did account for clustering using a multilevel model, using the same data as Chung et al. (2010) and found an odds ratio for any unit with admissions less than 100 to be higher than those with greater than 100 annual VLBW admissions. However Chung and colleagues, in their 2011 study, excluded some data from the earlier 2010 study and used a different classification system for unit volume and level making comparison of the results from the different methods difficult. The motivation for doing so was unclear.

Cifuentes et al. (2002) found that Californian infants treated in lower level units had an increased odds of mortality compared to larger units and that this effect was amplified when the data were restricted to VLBW infants or infants less than 1,250g. An earlier study by Menard et al. (1998) found the same result in South Carolina. Rogowski et al. (2004) found that in units with fewer than 50 VLBW admissions per year, an increase of 10 admissions per year was associated with an 11% reduction in mortality (95% CI: 5-16% p<0.001). Lower level units had higher mortality as did more urban ones. However, volume did not explain much of the variability in unit mortality (9%). Sanderson et al. (2000) and Samuelson et al. (2002) both found similar benefits for the largest units with the latter study estimating that 16-23% of deaths among VLBW infants could have been averted if 90% of VLBW births occurred in the hospitals with the highest level units. Bartels et al. (2006) examined the combined effect of size of NICU and the volume of the delivery unit on neonatal mortality, the results of which are shown in Figure 3.3.

Fig. 3.3 The estimated odds ratio for hospitals with large and small NICUs and large and small delivery units adjusted for a number of baby covariates, a large NICU was defined as having $\geq$36 VLBW admissions annually and a large delivery unit was defined as having $\geq$1,000 births a year, from Bartels et al. (2006)



All of the papers that were identified that looked at economic outcomes included some concept of volume. Four of the six economic outcome papers were from the UK, although all of these were published in 2000 or earlier.[16] The earliest paper was by Fordham et al. (1992) where the authors estimated a production function for neonatal care in the UK. The authors determined that intensive care was around three times more expensive than special care (these levels are now defunct) in terms of cost per cot day and the authors found that there are economies of scale which diminish until the unit has around 4,000 cot days annually (equivalent to approximately 13-14 cots). This led the authors to conclude that the cost-savings from centralisations may not be large enough to justify the extra travel required for parents to get to larger hospitals. In a response to this paper O'Neill and Largey (1997) reused the Fordham data and estimated a number of different functional forms for the production function and found quite different results depending on the functional form used. O'Neill and Largey (1997) did also conclude however that there were returns to scale in neonatal units which were dependent on the proportion of intensive care days provided by the units. Units were under increasing returns to scale as long as 64% or less of their care days

[16]Fordham et al. (1992); Hollingsworth and Parkin (2001); O'Neill and Largey (1997); O'Neill et al. (2000)

were intensive care; otherwise increases in volume led to increases in average costs. O'Neill et al. (2000) collected new data from English neonatal units and estimated a double log production function and concluded, again, that there were economies of scale present in English neonatal units. Specifically, the authors estimated that an increase from 2,000 to 3,000 cot days per year led to reductions in average costs of 19% and 16% for units with 5% and 10% of cot days as intensive care days respectively . The data from O'Neill et al. were re-used by Hollingsworth and Parkin (2001) to determine technical efficiency of English neonatal units. One of the finding of this study was that larger units were generally more efficient; however, specific figures were not provided for reasons unknown. The correlation between costs of care provision and unit volume is examined in Chapter 5.

Level of Care

Level of care was examined in 13 studies as a unit level variable.[17] Unit level is strongly related to unit volume as units with a wider range of facilities and greater resourcing tend to receive more infants from lower level units. Most (eight of thirteen) studies included separate volume and level indicators.[18] whereas Phibbs et al. (2007) and Phibbs et al. (1996) disaggregated the unit level variable into a number of categories, such as level 1 with fewer than or equal to 10 annual VLBW admissions, level 1 with greater than 10 VLBW admissions, level 2 with fewer than 10 VLBW admissions, and so forth. However, the definitions of volume used in the analyses alongside unit level as well as unit level definitions varied. Furthermore, for four studies volume was not included in the analyses as it was assumed to be different between unit levels. [19] This lack of uniformity in the analyses makes comparing outcomes from these studies difficult.

The estimated effect of unit level was generally of the same sign as volume; higher level units were found to have reduced mortality after case mix adjustment. Baker and

[17] Almond et al. (2010); Baker and Phibbs (2002); Bode et al. (2001); Chung et al. (2011, 2010); Cifuentes et al. (2002); Lorch et al. (2012); Phibbs et al. (2007, 1996); Rogowski et al. (2004); Shim et al. (2013); Straney et al. (2012); Wall et al. (2004)

[18] Almond et al. (2010); Bode et al. (2001); Chung et al. (2011, 2010); Lorch et al. (2012); Shim et al. (2013); Straney et al. (2012); Wall et al. (2004)

[19] Baker and Phibbs (2002); Cifuentes et al. (2002); Phibbs et al. (1996); Rogowski et al. (2004)

Phibbs (2002) estimated that there was a 38-45% reduction in odds of mortality for infants born in hospitals with high level NICUs compared to infants born in hospitals with no NICU. Cifuentes et al. (2002) looked at mortality for different groups of babies, grouped on the basis of birth weight. For the largest group, those born at less than 2,000g, the authors reported an odds ratio for mortality of 1.92 (95% CI 1.44-2.54) for care in intermediate NICUs versus regional NICUs.[20] The odds ratio decreased for small and large community NICUs; the estimated odds ratio for mortality for the latter was not significantly different from care in a regional NICU at the 5% level.

In his PhD thesis, Freedman (2010) examined the relationship between volume and level of care and neonatal outcomes using California data. The author attempted to demonstrate a causal relationship by using an instrumental variables (IV) approach. As an instrument, the author used the mother's distance to the nearest hospital, which exploits the strong, exogenous preference of an individual to go to their nearest hospital. Before using the IV approach, the author found a positive association between volume or level, and the risk of mortality such that higher volume and level units were associated with an increased risk of mortality; however, this effect disappeared when the IVs were used. Furthermore, the author found that the two stage least squares (2SLS; an estimator to calculate IV estimates) estimates were significantly different to the ordinary least squares (OLS) estimates. The author suggested that this may be due to unobservably higher risk mothers self-selecting into lower level hospitals so that relocating them to a higher level hospital would not necessarily reduce mortality. Similarly, Lorch et al. (2012) use an IV methodology to identify the effect of unit level on patient outcomes in three US states (Pennsylvania, California, and Missouri) and found a reduced risk of mortality for infants born in level three hospitals compared to their counterparts born elsewhere.[21] The authors found the magnitude of this effect differed by state, suggesting that other factors, such as labour or capital inputs, exacerbated or

[20]The classifications of neonatal units by Cifuentes et al. (2002) were as follows: No NICU—Cared only for healthy neonates and those with minor medical problems; Intermediate NICU—Cared for moderately sick infants but did not regularly provide assisted ventilation for more than 4 hours; Community NICU—Provided long-term ventilatory support but not all other specialized services normally provided by regional NICUs; Regional NICU—Provided a full range of specialized neonatal intensive care, including pediatric subspecialty consultants and surgery.

[21]Using the same definitions as Phibbs et al. (2007)

alleviated the effect of place of birth. This reflects the discussion in Chapter 1 where it was suggested that both the level of specific human capital embodied in the workforce, the level of capital inputs, and the way in which these combine all have implications for the quality of the healthcare provided.

Staffing

The effects of staffing were reported in 13 studies.[22] These include studies in which staffing outcomes were reported directly, for example the effect of an increase in the number or ratio of a certain type of staff, or indirectly, where the difference is examined between, for example, two times of day where staffing levels are different.

In a recent study, Lake et al. (2012) examined the difference in mortality and nosocomial infection for hospitals awarded a recognition for nursing excellence (RNE) in a cohort of 72,235 VLBW infants in the Vermont Oxford Network. The authors found a statistically significant (at the 5% level) reduction in the odds of early neonatal mortality (OR, 0.87; 95% CI, 0.76-0.99; p=.04), severe intraventricular haemorrhage (OR, 0.88; 95% CI, 0.77-1.00; p=.045), and infection (OR, 0.86; 95% CI, 0.75-0.99; p=.04). There was not a significant difference for neonatal or hospital stay mortality. Two studies from the UK also looked at the effect of nursing on outcomes. Tucker (2002) did not find a significant difference between hospitals with high and low nursing provision, although they did find, perhaps counter intuitively, that the odds of nosocomial bacteracaemia was lower in hospitals with low neonatal consultant provision (OR, 0.65; 95% CI 0.43-0.98). Conversely, Hamilton et al. (2007) found that an increase in the specialist nurse to patient ratio was associated with a reduction in risk-adjusted mortality; specialist nurses were defined as those nurses possessing qualifications from certain neonatal healthcare courses. Specifically, the odds ratio for mortality when the specialist nurse provision ratio was 1.3 to 1.8 was 0.52 (95% CI, 0.33-0.83) versus a comparative specialist nurse provision ratio <1.0. It should be noted that the 1.0-1.2 and >1.8 specialist nurse provision groups had non-significant odds ratio estimates.

[22] Abdel-Latif et al. (2006); Bell et al. (2010); Callaghan et al. (2003); Cimiotti et al. (2006b); Filho et al. (2011); Goodman et al. (2002); Grandi et al. (2010); Hamilton et al. (2007); Lake et al. (2012); Profit et al. (2010); Straney et al. (2012); Synnes et al. (2006); Tucker (2002)

Both of these UK based studies used the same data (from the UK Neonatal Staffing Survey (UKNNSS), details of which are published in Tucker (2002)); Hamilton et al. (2007) examined only VLBW infants whereas Tucker (2002) examined all admissions. However, both of these studies used a cross-classified statistical model, which simultaneously examines the effects of high/low nursing provision,[23] the effect of unit volume, and the effect of neonatal consultant provision. This makes interpretation of the effects of nurse provision difficult, since it cannot necessarily be disentangled from the effects of the other variables that are examined.

The most recent of the previously listed studies used a sample of moderately preterm infants (born between 30^{+0} and 34^{+6} weeks gestation) and examined the relationship between the mean nurse to patient ratio and a range of clinical outcomes (not including mortality) (Profit et al., 2010). The only statistically significant (at the 5% level) effect observed from a series of linear regressions was that an extra patient per nurse led to a decrease in average daily weight gain by 24%. None of the studies identified in the systematic literature review attempted to estimate a causal effect. The associations reported in the studies are likely to be subject to bias owing to the system of transfers that mean the location and characteristics of the unit providing treatment are determined by the infant's healthcare needs. Similar to the aforementioned studies, Filho et al. (2011) found an association between nurse to patient ratios and adverse events associated with mechanical ventilation.

Three studies looked at the effect of supply of neonatologists on outcomes.[24] Synnes et al. (2006) looked at the effect of a number of unit characteristics, including neonatologist/household staff ratios and patient volume, on the incidence of severe IVH. The authors found that units with a higher neonatologist to housestaff ratio had a lower odds ratio for severe IVH (OR, 0.2; 95% credible interval[25] 0.0-0.7). Goodman et al. (2002) studied the difference in neonatal mortality between different regions and found that, after adjusting for confounding baby and mother covariates, areas with 4.3 neonatologists per 10,000 births (the second lowest quintile) were associated with a reduction

[23]Defined as having greater/less than 0.85 nurse to cot ratio (the median) in both studies.
[24]Goodman et al. (2002); Straney et al. (2012); Synnes et al. (2006)
[25]A credible interval is the Bayesian equivalent of the frequentist confidence interval.

in the odds of mortality when compared to areas with 2.7 neonatologists per 10,000 births (the lowest quintile) (OR, 0.93; 95% CI, 0.88-0.99), although further increases in neonatologist supply were not associated with reductions in mortality. Finally, Straney et al. (2012) found that an increased supply of neonatologists was associated with a reduction in the preterm adjusted mortality rate.

The studies that looked at staff provision indirectly did not find a difference in mortality between babies admitted in the evening or at weekends in Australia (Abdel-Latif et al., 2006; Bell et al., 2010), but did find that the risk of stage 2 or higher retinopathy of prematurity (ROP) was greater after the introduction of duty hour restrictions for residents in neonatal units in Iowa, US. Sanderson et al. (2000) reported that units with 24-hour neonatology cover had shorter lengths of stay and lower Medicaid reimbursement, although, oddly, they do not report any estimates for the effects.

Other Unit Characteristics

Other unit characteristics were used in a number of studies, including the presence of neonatal fellowship or consultancy positions (Chung et al., 2011) or level of revenue (Wall et al., 2004). One study looked at the effect of organisational characteristics of neonatal unit, such as leadership and coordination, which were determined by staff completed questionnaires, on a wide range of clinical outcomes (Pollack and Koch, 2003). The authors reported that a lower incidence of peri-intraventricular haemorrhage (PIVH) or periventricular leukomalacia (PVL) was associated with better scores for leadership, coordination, and conflict resolution and better values in the therapists', nurses', and physicians' scores were associated with lower mortality, bronchopulmonary dyplasia (BPD), IVH or PVL, and ROP (Pollack and Koch, 2003). Although, as the authors concede, given the abstract nature of the organisational characteristics, like leadership, no specific problem was isolated at which an intervention could be directed to improve neonatal care. These characteristics may be correlated with both the specific and general human capital in the unit's labour force. Identification of the relationship between these factors and other observable characteristics of the neonatal unit may advance our understanding of human capital development within neonatal

units.

One study, of particular importance to the theme of this thesis and the only one identified to have used data from the NNRD, examined the location of birth and probability of transfer before and after the reorganisation of neonatal units into managed clinical networks in England (Gale et al., 2012b). The authors found that when comparing infants born at 27-28 weeks gestation between September 1998 and August 2000 and those born between January 2009 and December 2010 that:

> there were increases in the proportions of babies born at 27-28 weeks' gestation in hospitals providing the highest volume of neonatal specialist care (18% (631/3495) v 49% (1325/2724); odds ratio 4.30, 95% confidence interval 3.83 to 4.82; P<0.001) and in acute and late postnatal transfers (7% (235) v 12% (360) and 18% (579) v 22% (640), respectively; P<0.001). (Gale et al., 2012b)

3.2.2 Research Methods

The methods and results of individual studies discussed in this section are provided in Appendix A.

In a general sense, the overarching aim of the empirical chapters in this thesis is to identify the causal effects of specific unit characteristics, such as unit volume or a nurse to patient ratio, on the clinical outcomes of infants treated within that unit. To be able to estimate these effects, which could be considered treatment effects, a counterfactual is required—information on what would have happened to the infant both with and without the treatment. However, empirical counterfactuals do not exist for observational data in general and so econometricians often rely on exploiting a quasi-experimental design—i.e. using exogenous variation in the treatment of interest. Typically, in studies of medical interventions, randomised controlled trials are used to estimate a treatment effect; the experimental design allows the researcher to control for many of the variables that may also affect the outcome. Observational studies on the other hand do not benefit from an experimental design. Infants are not randomised

to hospitals. Indeed the lack of randomisation is caused, in part, by the fact that infants can be transferred both in utero and post-natally to appropriate units.

Econometrics provides tools, such as regression, with which to analyse observational data. As the literature review showed, the vast majority of previous studies in this area utilised some form of logistic regression (for example, Chung et al. (2010); Phibbs et al. (2007)). A benefit of regression techniques is that they allow the researcher to 'hold fixed' many confounding variables to investigate the effect of a change in the variable of interest. This may permit conclusions of the form: infants born in hospitals with a greater annual average number of admissions have a lower risk of in hospital mortality, *ceteris paribus*. However, there are a number of issues that may preclude accurate inferences when determining the effects of unit level variables on individual outcomes.

A variable is said to be endogenous if it is correlated with the error term in a regression.[26] Endogeneity is a concern, particularly if the researcher is interested in causal inference, as this implies that certain estimators are biased. Endogeneity is likely to be present in regression models of neonatal unit level variables on individual outcomes, principally due to the fact that babies are transferred between units. If the dependent variable is a health outcome such as mortality, and there is heterogeneity between infants that is unobservable to the analyst, and if this unobserved heterogeneity determines in part the location of birth, then the unit characteristics are endogneous in our statistical models. Broadly speaking, this endogeneity arises since there is a loop of causality between latent health and the characteristics of the neonatal unit providing treatment. Furthermore, it is possible that mothers may self-select into different hospitals on the basis of unobserved characteristics as well.

None of the previous studies discussed in the literature review addressed this problem of endogeneity with the exception of Lorch et al. (2012), Almond et al. (2010), and Freedman (2010). These results were discussed in the previous section. As a result,

[26]Endogeneity may arise for one of three reasons: (i) omitted relevant variables, a variable correlated with both the treatment and outcome is omitted from the experimental design; (ii) simultaneity, the outcome has a causal effect on the treatment; and (iii) measurement error, the treatment status or its value may be measured with error.

none of the previous studies, except these three studies, can reasonably argue to have attempted to estimate causal effects.

The most frequently utilised solution to endogeneity bias is to use an instrumental variables approach, although other techniques are available. An instrumental variable (IV) is one which is correlated with the endogenous regressor but not the error term such that it is a source of exogenous variation in the endogenous regressor of interest. An IV is able to recover the local average treatment effect (LATE) if the IV meets certain assumptions (Imbens and Angrist, 1994). The LATE is the average treatment effect for the subpopulation of individuals whose treatment status is altered by the instrument. This point is important for interpretation later.

The majority of previous studies identified a population averaged difference between neonatal unit characteristics as opposed to an individual level effect as all observations from all hospitals were pooled into one dataset. However, in many cases the observations within a neonatal unit or hospital are correlated with one another. A multilevel model is one that can be applied to data with a nested structure so that observations from the same area or unit may be correlated; for example, one would expect that the outcomes of babies in the same neonatal unit would be correlated. The multilevel model (in a two level case) specifies the coefficients in the level one model (in this case the individual baby and the intercept) to be determined by a linear function of a random term and level two variables (here, the hospital unit characteristics). This model can be simplified to only allow the intercept term to vary by cluster (a varying intercept model) in which case the model is analogous to a panel data model with an unobserved individual heterogeneity term, but instead of multiple observations of one individual over time there are multiple observations of one cluster over a cross section. This specification then allows the researcher to estimate individual level effects; only three studies identified in the literature review used such a method (Chung et al., 2010; Lake et al., 2012; Synnes et al., 2006) and in each case the simpler varying intercept model was used.

As the literature review identified, a number of outcomes have previously been studied. These are predominantly clinical, notably mortality (e.g. Chung et al. (2010);

Phibbs et al. (2007)), but also included intraventricular haemorrhage (IVH) and nosocomial infection (e.g. Pollack and Koch (2003); Synnes et al. (2006)). Generally these studies have therefore had a binary dependent variable and utilised a logistic regression. However, the use of more complex models may be required in the case of a multinomial outcome and/or if duration is of interest. The former case may arise, for example, if one were interested only in the outcome of a first episode of care when the possible outcomes are mortality, discharge from neonatal healthcare, or transfer to a different neonatal unit.

In the multinomial case, the logit model is easily extended to accommodate the different outcomes leading to, for alternative invariant regressors, the multinomial logit. However, a general weakness of these types of models is the independence of irrelevant alternatives (IIA) assumption, which means that the conditional probability of one outcome compared to another is independent of the other outcomes. This assumption is a strong one and in the case of first episode outcomes, where the potential outcomes are mortality, discharge, or transfer, wrong since if transfer by another method of transportation became a possible outcome of care, then we would not expect the probability of mortality or discharge to change and the probability of experiencing the original form of transport to halve, which violates the IIA assumption. The IIA assumption may be relaxed; if the data display a nested structure, then a nested logit may be used, although in the earlier example this is not the case. Alternatively, a random parameters logit may be used which specifies the parameters to be random variables and thus permits correlation across alternatives. These types of models may also be useful when examining the destination of transfers. Furthermore these types of model also permit the analysis of ordered or sequential outcomes, which may be highly useful for morbidities with ranked degrees of severity such as retinopathy of prematurity.

Duration may also be an important consideration in models that examine length of stay or the duration of intensive care, for example. In this case, survival analysis techniques are needed. In many instances of survival analysis the event of interest is only witnessed as occurring in a certain interval and so a discrete time formulation is typically used. However the data in the NNRD includes the time that each key event in

an infant's healthcare occurred at minutes past birth meaning continuous hazard functions could be used. Censoring may also have to be accounted for, although only a small number of babies are transferred to hospitals not contributing data to the NNRD. The most popular models in the survival analysis literature are proportional hazards (PH) models of which there are continuous time and discrete time variants. The most well-known PH model is the Cox PH model, a semi-parametric specification that does not specify an explicit baseline hazard. Fully parametric models have a lot of use, but suffer from inconsistent parameter estimates if the baseline hazard is misspecified; a flexible baseline hazard may be used although this approach may lead to difficulties with estimation and with identification. In a PH specification the covariates are modelled as having a scaling effect on the baseline hazard which assumes that they are time invariant, this may be relaxed by including interactions with time in either a continuous or discrete time setting.

In the case where there are multivariate duration data, a number of frameworks are available of which the competing risks (CR) framework is perhaps the most popular. This extends some of the survival models mentioned above to jointly model hazard functions for different outcomes. This approach is required when different hazards may be linked such as with different morbidities or first episode outcomes. Alternatively the goal may be to jointly model two different outcomes such as duration of intensive care or a central line and time to a particular comorbidity such as infection. The CR framework can incorporate dependent risks by allowing an unobserved heterogeneity term to be correlated between outcomes, furthermore this approach can be used in the PH framework. While these types of model are very rich they are computationally very difficult; the maximum likelihood estimates require the solution of a high dimensional integral which does not have an analytic expression, which then requires methods such as Monte Carlo integration. The large amount of data then makes computation of these estimates even more difficult.

A final issue that is relevant to this analysis is the attribution of outcomes. Since many babies are transferred during the course of their care, there will be more than one unit that the unit level variables can be taken from. The majority of studies identified

in the literature review focussed on place of birth (e.g. Cifuentes et al. (2002); Phibbs et al. (2007)). Thus their inference was only about the effects of the unit in which a baby was born rather than the unit providing treatment even if the two may be considered to be strongly correlated. Ideally, one would want to determine the effects of the unit providing care and at each point during the care process. Tucker (2002) assigned the baby to the unit the baby was in at 24 hours post birth, this allowed for any transfers within the first day of life. The authors tested the robustness of their results to using place of birth and found little difference.

3.3 Discussion

Overall, there is a fairly rich literature examining the relationship between neonatal unit characteristics and outcomes, however these are subject to a number of methodological concerns as highlighted in this chapter. There is a fairly strong consensus among US studies that high level, higher volume units are associated with reduced mortality and morbidity, although this relationship has not been demonstrated in the UK. Moreover, as has been discussed earlier in this chapter, there are a large number of differences between the US and UK health care systems preventing comparisons of the two bodies of literature. It is possible that patients self-select to better hospitals in the US due to greater patient choice (at least among certain groups of patients). Those mothers with enhanced insurance, who are likely to be better off, are more likely to be the mothers that have healthier babies. It would not be surprising then if the healthier babies went to the hospitals with better outcomes anyway. This could in part explain why there are high risk babies being treated in low level hospitals in California. Indeed, Freedman (2010) does suggest this to be the case.

There are very few studies that have looked directly at the effects of neonatal unit characteristics from Europe and the UK. More research from the UK is clearly warranted, particularly since a key study, Tucker (2002), was published in 2002, before the reorganisation of neonatal units into networks. Indeed, the most recent data to have been used in the studies reviewed here on volume used data for babies born in 2005

(Lorch et al., 2012).

Another point of interest is that those studies which report the proportion of variance attributable to unit level characteristics indicate that there is much greater variability among morbidity outcomes than mortality. This should not be unexpected since technologies that prevent mortality, such as surfactant therapy, tend to be rapidly and efficiently disseminated among units. However, for many neonatal morbidity outcomes the aetiology is not well understood and so there are few technologies that are used to prevent a morbidity outcome. For morbidity outcomes there may be a large role played by hygiene, feeding practices, or physician skill, all of which vary from unit to unit.

The majority of studies reviewed here focussed on only VLBW babies, or another subgroup of infants with a high rate of mortality. There was very little research based on a more representative cohort of infants. It is unclear why this may be. It is possible that the data were unavailable for these infants. Another explanation may be because of a possible selection effect for healthier babies—not all healthier babies are admitted to neonatal care and so there may be an unobservable selection mechanism that depends on the unit. Indeed, as hitherto discussed, one study found that the size of the delivery unit affected the likelihood of admission to a neonatal unit (Le Ray et al., 2009). However, none of the studies identified in this review stated this. Furthermore, a similar line of reasoning could be argued for very preterm babies, all of whom are admitted—the skill of the delivery suite determines which babies survive and are admitted to neonatal care.

Studies examining the relationship between volume and outcomes in other areas of medicine generally focus on individual procedures. Surgical procedures have often been the subject of volume-outcome analysis. A recent example shows a positive effect of volume of procedures on outcomes of cholecystectomy (Harrison et al., 2012). Neonatal care is different as it comprises a wide range of procedures entailing the full care pathway of an infant. Intensive care is one important aspect but so is a ‘feed’ and ‘grow’ regimen for healthier babies. In order to consider volume as a variable affecting outcomes this should be taken into account by considering different procedures separately and together. The following chapter estimates the effect of neonatal unit

designation and volume at the hospital of birth on the risk of mortality and a number of morbidity outcomes.

Chapter 4

The effect of volume and designation of care on neonatal clinical outcomes

The work presented in this chapter has since been published as:

> Watson, S. I., Arulampalam, W., Petrou, S., Marlow, N., Morgan, a. S., Draper, E. S., and Modi, N. (2014). The effects of designation and volume of neonatal care on mortality and morbidity outcomes of very preterm infants in England: retrospective population-based cohort study. BMJ Open, 4(7), e004856–e004856. doi:10.1136/bmjopen-2014-004856.

This chapter represents work conducted as part of a collaborative research project. The collaborators and contributions to this work are presented in the Declaration.

4.1 Introduction

Intense debate, both in the United Kingdom and internationally, has revolved around the effect of the volume and designation of the neonatal unit at the hospital of birth on the clinical outcomes of infants admitted to these units (see Chapter 3 for a full review). This debate has important ramifications for the optimal design of neonatal critical care services at the national level. As detailed in Chapter 3, a large number of studies have suggested that the intensity and volume of neonatal care at the hospital of birth is negatively correlated with adverse clinical outcomes, including mortality (Bartels et al.,

2006; Chung et al., 2011, 2010; Cifuentes et al., 2002; Fellman et al., 2009; Johansson et al., 2004; Lasswell et al., 2010; Lorch et al., 2012; Phibbs et al., 2007; Rautava et al., 2007; Rogowski et al., 2004; Synnes et al., 2006), intraventricular haemorrhage (Synnes et al., 2006), infection (Lorch et al., 2012), and bronchopulmonary dysplasia (Lorch et al., 2012).

All of the previous studies have been conducted in countries without, or prior to the formation of, a managed clinical network (MCN) system. The MCN system for neonatal care was designed to replicate the benefits of centralisation without sacrificing equity of access to neonatal care through a dedicated system of inter-unit transfers (Department of Health, 2003). Recent research has shown that since the formation of MCNs in 2006 in England, both the proportion of infants born at 27-28 weeks gestation born in hospitals with tertiary level neonatal units, and the proportion of those infants receiving a transfer has increased (Gale et al., 2012b). The aim of this chapter is to determine whether MCNs have achieved one of their goals or providing the benefits of centralisation by examining the effect of designation and volume of care of the neonatal unit at the hospital of birth on mortality and morbidity outcomes.

Both designation and volume have been the subject of numerous studies which generally find a reduction in the risk of mortality and morbidity for infants born in hospitals with the largest and/or highest designation neonatal units. However, volume and designation are highly correlated with the largest units having the most intensive facilities. The volume-outcome relationship has been documented in a number of healthcare fields (Halm et al., 2002b; Luft et al., 1979, 1987). These studies have had a large impact on healthcare policy, with many authors advocating a centralisation of healthcare services. However, from a policy perspective, identifying the direction of causality in the volume-outcome relationship is crucial for making these decisions appropriately. Only a handful of studies out of the hundreds published have attempted to identify a causal effect (Barker et al., 2011; Halm et al., 2002b). The two principle hypotheses about how the volume-outcome relationship functions are ‘practice makes perfect’, which describes the causal effect of volume on outcomes, and ‘selective referral’, which describes the causal effect of outcomes on volume (Luft et al., 1987). The

latter case arises when hospitals with superior outcomes attract a greater demand for their services such as through a dedicated inter-hospital transfer system. The former hypothesis may be decomposed into the effects of economies of scale and the effects of learning by doing. Economies of scale could influence quality per se by leading to an increase in labour and capital inputs. Learning by doing increases the specific human capital embodied in the labour force which in turn may increase the efficiency with which capital inputs are utilised in the production of healthcare. Moreover, the marginal benefit of increased specific human capital may be greater for more complex cases, suggesting that there may be a 'steeper' learning curve for more complex cases. However, the causal mechanism by which designation may affect outcomes above and beyond those of volume or greater resourcing are unclear.

The most widely used method to estimate the effects of the characteristics of the place of birth on clinical outcomes is an adjusted regression that takes account of various exogenous clinical characteristics of the patient that may determine outcomes (Halm et al., 2002b). However, this method does not allow for a correlation between the place of birth and unobserved patient heterogeneity caused by 'selective referral'. This simultaneity is clearly the case in the MCN system. Not accounting for such simultaneity between health and place of birth may lead to biased estimators of the effect of interest. In this chapter, an instrumental variable (IV) methodology is used to attempt to identify the causal effect of unit volume and designation on clinical outcomes. Only one previous study to examine the effect of the volume of the neonatal unit at the place of birth on mortality outcomes in neonates has utilised an IV method (Lorch et al., 2012). Examining all hospital based deliveries in Pennsylvania and California between 1995 to 2005 and Missouri between 1995 and 2003 with a gestational age between 23 and 37 weeks gestation, Lorch et al. (2012) used a matched-pair instrumental variables design and exploited the variation in delivery hospital caused by differing travel times to various types of hospital (the methodology is explained in detail by Baiocchi et al. (2010)). The authors found that:

> Infants who were delivered at a high-level NICU had signifi- cantly fewer

> in-hospital deaths in Pennsylvania (7.8 fewer deaths/1000 deliveries, 95% confidence interval (CI) 4.1–11.5), California (2.7 fewer deaths/1000 deliveries, 95% CI 0.9–4.5), and Missouri (12.6 fewer deaths/1000 deliveries, 95% CI 2.6–22.6). (Lorch et al., 2012)

The rest of this chapter is organised as follows. Section 4.2 discusses the sample of infants selected for this study, Section 4.3 provides the definitions of volume and designation used, and Section 4.4 explains the econometric method utilised. The results are provided in Section 4.5 along with robustness tests in Section 4.6, and Section 4.7 concludes.

4.2 Sample selection

From the NNRD, data are extracted on all infants born at $\leq 32^{+6}$ weeks$^{+\text{days}}$ gestation between January 1$^{\text{st}}$ 2009 and December 31$^{\text{st}}$ 2011. All infants born below this gestational age threshold are classified as very preterm; this is a frequently studied patient group, Table 2.3 in Chapter 2 provides definitions of gestation age groups used in studies of neonates. This gestational age cut-off is selected for a number of reasons: firstly, as outlined in chapter Chapter 2, above this cut-off very few infants experience the clinical outcomes used in this part of the study; secondly, the function relating observed covariates to outcomes may be different between infants either side of this cut-off—as is shown in Figure 2.3, Chapter 2; thirdly, analysing only very preterm infants enables greater comparability between these results and those of other related studies which generally only focus on this group of infants; and finally, not all infants born above this threshold are admitted to neonatal units, so there may be a sample selection issue in larger units.

The baseline analysis examines infants born at $\leq 32^{+6}$ weeks gestation. Further analyses are conducted on two sub-samples of infants, those born at $\leq 26^{+6}$ weeks gestation and those born at $27^{+0} - 32^{+6}$ weeks gestation. The reason for separating these sub-samples is that most neonatal networks aim to transfer all women at high risk of delivery at <27 weeks gestation prenatally to hospitals with level three neonatal

units. In addition, since most relevant studies based in the United States examine only VLBW infants, we also examined this group for further comparability between studies. Infants who received only transitional care (defined by HRG code 'XA04Z') are excluded from the analysis (n=5).

4.3 Variables

4.3.1 Volume and Designation of Care

The volume and designation variables are both coded as binary indicators (see section 4.4), they can therefore be viewed them as binary treatments. In order to dichotomise volume and designation, the following definitions are used where the treatment indicator was equal to one for an infant if the infant is:

- **Tertiary level**: Admitted to a tertiary neonatal unit at the hospital of birth.
- **High volume**: Admitted to a neonatal unit in the top quartile of neonatal units by volume at the hospital of birth.

For the definition of 'high volume', the primary measure of volume was the annual number of care days at any level of care provided to very preterm infants. The results are examined for robustness to alternative definitions of volume, in particular, the annual number of very preterm births and admissions to the unit, and the annual number of intensive care days provided. 'High volume' is defined by quartile rather than an absolute care day threshold to facilitate comparison with other measures of volume in the sensitivity analyses. A previous study that examined organisational characteristics of neonatal units also categorised volume using quartiles (Van Reempts et al., 2007). In addition, we also define 'high volume' as greater than 100 VLBW admissions in a year to compare the results to Phibbs et al. (2007) who use the same definition.

These definitions state that the treatments include only those infants who are 'admitted to... at the hospital of birth' for a certain type of unit rather than those simply 'born in a hospital' with a certain type of unit. This is to make it clear that only neonatal admissions are observed in the NNRD—infants who are born in a hospital but die

prior to admission are not observed. This is a possible weakness of this study, the implications of which are further discussed in Chapter 9. However, if anything, not observing these infants is likely to lead us to underestimate the benefits of birth in a hospital with a high volume neonatal unit.

4.3.2 Outcomes

The following outcomes are considered in this chapter, all of which are binary and are defined and described in greater detail in Chapter 2:

- Neonatal mortality
- Any in-hospital mortality
- Bronchopulmonary dysplasia
- Surgery for necrotising enterocolitis
- Treatment for retinopathy of prematurity
- Discharge after 40 weeks post menstrual age.

The final one of these outcomes is used to capture a long length of stay that may both reflect poor health and economic burden.

The morbidity outcomes used in this chapter are frequently used in studies of this nature. However, as was discussed in Chapter 3, standard estimators of models with these morbidities as outcomes may be inconsistent due to the fact that many infants who would be likely to experience the morbidity outcome of interest die prior to being observed with the morbidity. This is censoring due to mortality. An alternative model that incorporates multiple outcomes may be preferred in this case. However, for the purposes of this chapter, we instead perform two sensitivity checks: first, we re-conduct the morbidity analysis not including infants who died in the sample, and secondly, we define a new outcome for whether an infant experienced the morbidity and/or died.

4.3.3 Covariates

Chapter 2 provides an overview of the NNRD and the availability, quality, and summarises the available variables. Following a review of previous prediction models for very preterm infants (Medlock et al., 2011), covariates are selected that a) are significant predictors of adverse sequelae, b) are available in our dataset and of high quality, and c) not confounded by the provision of neonatal care. The variables included are: gestational age at birth, gestational age squared, birth weight z-score (birth weight normalised by gestational age week), and the following indicators: whether the mother received a full or partial course of antenatal steroids, male sex, infant year of birth, and whether or not the mother came from an area within the lowest decile of the Index of Multiple Deprivation 2007 score (see Chapter 2, Section 2.3.2 and Noble et al. (2007)).

4.3.4 Instruments

The system of transfers in England and Wales may lead to a correlation between infant health, both unobservable and observable, and the volume and designation of the neonatal unit at the hospital of birth—infants are typically transferred to units designated to provide the appropriate level of care; volume is strongly correlated with unit designation (Figure 2.2 in Chapter 2). Instrumental variables for the volume and designation of the neonatal unit at the hospital of birth are therefore required. The conditions that these instrumental variables must satisfy are detailed in the following section, section 4.4. For the instruments, we use indicators for the designated level of care of the nearest neonatal unit to the mother's residence, an indicator for whether it had surgical facilities, an indicator for whether it was high volume, the distance to the nearest neonatal unit, and the interactions of either the level of care indicators or high volume indicator with distance, giving nine instruments in total. Straight line distance is calculated from the population weighted centre of the mother's Lower Super Output Area to each hospital (Office for National Statistics, 2011). Geographical variation in distance to certain institutions as a source of exogenous variation in use of or access to these institutions has been widely used in studies of both healthcare and other ar-

eas. For example, in a widely cited study, Card (1995) used proximity to college as an exogenous determinant of schooling since it was found that men who grew up in local labour markets with a nearby college has significantly higher education than other men.

4.4 Econometric specification

Two separate sets of analyses are conducted based on whether or not infants were admitted to a i) tertiary level, or ii) high volume neonatal unit (defined in section 4.3.1) at the hospital of birth. For these analyses an appropriate statistical model is required, see Chapter 3 for a review of methods used in the literature previously. All of the outcome variables are binary and are thus Bernoulli distributed, however, the distribution of the probability of the outcome, p, is unknown. Typical models for p are the logit model, $p = \Lambda(z) = 1/(1+exp^{-z})$ where $z = x'\beta$ and x are the observed covariates; the probit model $p = \Phi(z)$ where Φ is the standard normal cumulative distribution function; and, the linear probability model (lpm), $p = z$. The choice of the model for p should reflect the data generating process (dgp) as well as possible. While the dgp is unknown, we can aim for the 'best' approximation to it. The lpm is unlikely to be the best approximation to the dgp since p is not constrained between zero and one. However, the lpm may be viewed as a non-parametric version of the binary choice model where we assume linear conditional expectations. Nonetheless, unless all predicted probabilities from the lpm are constrained between zero and one then ordinary least squares estimators of the lpm have been shown to be both biased and inconsistent (Horrace and Oaxaca, 2006).

In choosing between logit and probit models, the theoretical consequences of model misspecification are not great since the ratio of the slope parameters between models is constant if the regressors are distributed so that the condition mean of each regressor is linear in $x'\beta$ (Ruud, 1983). Probit models are often utilised since they are motivated by a latent normal random variable. However, in this case, the logit model was used to remain consistent with the previous literature in this field (for example, Chung et al.

(2010); Phibbs et al. (2007)).

The analyses are conducted in two parts: firstly, a 'standard', adjusted logistic regression not accounting for infant unobserved heterogeneity; and secondly, an instrumental variable (IV) logistic regression.

These 'standard' and IV logistic regressions are as follows. Here, the exposition of Terza et al. (2008) and Wooldridge (2003) is used. Let y_i be the binary outcome for infant i, D_i be the tertiary level/high volume treatment indicator, let x_i be a vector of observed, exogenous covariates including an intercept, and let c_i be a scalar representing unobserved determinants of mortality correlated with D_i:

$$y_i = \Lambda(\delta D_i + x_i'\beta + c_i) + u_i \tag{4.1}$$

where u_i is a random error term with zero conditional mean, so that the conditional mean of y_i is:

$$E(y_i|x_i, D_i, c_i) = Pr(y_i = 1|x_i, D_i, c_i) = \Lambda(\delta D_i + x_i'\beta + c_i) \tag{4.2}$$

where δ and β are parameters to be estimated. The parameter of interest in this chapter is δ.

As previously mentioned, there are two parts to the analysis: a 'standard' adjusted logistic regression, and an IV logistic regression. The 'standard' logistic regression assumes that the population distribution of unobserved heterogeneity has zero variance such that there are no unobserved differences between infants, in which case c_i would be absorbed into the intercept. For the IV logistic regression, the relationship between the endogenous variable and other variables is modelled as:

$$D_i = \Lambda(x_i'\pi_1 + z_i'\pi_2) + v_i \tag{4.3}$$

where v_i is a random error term with zero conditional mean, so that the conditional

mean of the treatment is:

$$E(D_i|x_i,z_i) = Pr(D_i = 1|x_i,z_i) = \Lambda(x_i'\pi_1 + z_i'\pi_2) \tag{4.4}$$

where z_i is a vector of IVs (described in section 4.3), and $\pi = [\pi_1, \pi_2]'$ is a vector of reduced form parameters to be estimated. In addition, assume that (c_i, v_i) is independent of z_i and that:

$$c_i = \rho v_i + e_i \tag{4.5}$$

where e_i is independent of v_i (and necessarily of D_i). Under these assumptions, we have

$$E(y_i|x_i,D_i,v_i) = Pr(y_i = 1|x_i,D_i,v_i) = \Lambda(\delta D_i + x_i'\beta + \rho v_i). \tag{4.6}$$

Since v_i is unobserved, it is replaced in equation (4.6) with estimates. Let $\hat{\pi}$ be the estimates of the parameters in the first stage equation (4.4). The generalised residuals are then obtained as $\hat{v}_i = D_i - \Lambda(x_i'\hat{\pi}_1 + z_i'\hat{\pi}_2)$, and the regression model (4.6) is estimated with $(x_i, D_i, \hat{v}_i)$ as regressors. This method is also referred to as two stage residual inclusion (2SRI) (Terza et al., 2008; Wooldridge, 2003).

The instrumental variables must satisfy three conditions to identify the local average treatment effect (LATE) (Imbens and Angrist, 1994):

- Independence: the instruments are as good as randomly assigned, conditional on the observed covariates, so that $Cov(z_i, c_i|x_i) = 0$.
- Exclusion restriction: The instruments act only through the endogenous variable and do not have a direct effect on infant health. This is equivalent to saying that z_i does not feature in (4.1).
- Non-zero effect: the location of birth is a non-trivial function of the instruments. Specifically, we require $\pi_2 \neq 0$ in equation (4.3).
- Monotonicity: The instrument only shifts infants into the treatment and does not shift people out of the treatment. This can also be interpreted as there being no

defiers.[1]

These assumptions are further discussed in the following section and in the next chapter in the context of the model presented there.

The majority of the literature related to the analyses in this chapter is published for a clinical audience. The standard in the clinical literature is to present odds ratios (OR) for treatments, these are calculated as $OR = exp(\delta)$ for equation (4.1). For the primary analyses, I present ORs for comparability with the previous literature. However, I also provide estimates of average partial effects (APEs) for the main results as this is typical within the economics literature:

$$APE = E_{c_i}\left(\frac{\partial Pr(y_i = 1|x_i, D_i, c_i)}{\partial D_i}\right) \tag{4.7}$$

which is the partial effect of the treatment averaged across the distribution of the unobserved heterogeneity.

The analyses are conducted in R 2.15.3 and Stata version 11.

4.5 Results

4.5.1 Summary statistics

In total, the sample contains data from 20,554 infants born at $\leq 32^{+6}$ weeks gestation over the period 2009-11, 2,559 of whom were born at $\leq 26^{+6}$ weeks gestation. Table 4.1 provides descriptive statistics of the samples analysed. Overall, 9,466 (46.1%) infants were born in hospitals with a tertiary level neonatal unit and 9,541 (46.4%) were

[1]Following Imbens and Angrist (1994), there are four types of individuals with respect to our treatment (high or low volume place of birth, for example) and instrumental variable (high or low volume nearest neonatal unit, for example): *Compliers*: mothers who give birth in the nearest hospital regardless of whether it has a high level unit or not—if a mother lives near a high (low) level unit, then she gives birth in the hospital with a high (low) level unit. *Always-takers*: mothers who always go to a hospital with a high level or high volume unit. This could be mothers who have been assessed to be better off having the baby in a high level unit and they go there regardless of the distance. *Never-takers*: mothers who always go to a hospital with a low level or low volume unit. This could be because there is a policy that all mothers are taken to a low level unit without taking the risk into account and then infants are transferred after birth. This is unlikely and as such there are unlikely to be never-takers. *Defiers*: women who do the opposite of compliers. There are unlikely to be mothers that fall into this group.

Table 4.1 Summary statistics of the sample

	Designation			Volume[a]		
	Tertiary level unit	Non-tertiary level unit	p-value[b]	High volume unit	Non-high volume unit	p-value[b]
n (%)	9,466 (46.1)	11,088 (54.0)		9,541 (46.4)	11,013 (53.6)	
Gestation (weeks), mean (SD)	29.2 (2.5)	30.0 (2.1)	<0.001	29.3 (2.5)	29.9 (2.2)	<0.001
Birth weight (g), mean (SD)	1,313.9 (438.7)	1,451.9 (404.5)	<0.001	1326.6 (436.7)	1441.8 (409.4)	<0.001
Antenatal steroids[c]	6,394 (67.6)	7,262 (65.5)	0.002	6,330 (66.4)	7,326 (66.5)	0.790
Deprivation score bottom 10%[d]	2,020 (21.4)	1,342 (12.1)	<0.001	1,730 (18.1)	1,632 (14.8)	<0.001
Male	5,048 (53.3)	5,397 (53.4)	0.756	5,093 (53.4)	5,892 (53.5)	0.863
Neonatal mortality	423 (4.5)	366 (3.3)	<0.001	394 (4.1)	395 (3.6)	0.043
Any in-hospital mortality	569 (6.0)	425 (3.8)	<0.001	527 (5.5)	467 (4.2)	<0.001
BPD	3,695 (39.0)	2,856 (25.8)	<0.001	3,548 (37.2)	3,003 (27.3)	<0.001
Treatment for ROP	226 (2.4)	107 (1.0)	<0.001	195 (2.0)	138 (1.3)	<0.001
Surgery for NEC	167 (1.8)	123 (1.1)	<0.001	163 (1.7)	127 (1.2)	0.001
PMA at discharge >40^{+0} weeks	1,292 (13.7)	848 (7.7)	<0.001	1,237 (13.0)	903 (8.2)	<0.001

[1] All values are n (%) and are a proportion of the column total unless otherwise stated.
[2] BPD = Bronchopulmonary dyplasia, ROP = retinopathy of prematurity, NEC = necrotising enterocolitis.
[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.
[b] Continuous variables were tested by t-test, categorical variables by chi-squared test.
[c] Whether the mother received a full or partial course of antenatal steroids.
[d] Whether the mother's residence was in one of the 10% most deprived LSOAs measured by the Index of Multiple Deprivation.

born in hospitals with a high volume neonatal unit. The top quartile of volume was defined as approximately 3,480 annual care days for infants born at $\leq 32^{+6}$ weeks gestation. 165 hospitals are represented in the sample, 44 (26.7%) of which had level three neonatal units, 81 (49.0%) level two neonatal units, and 39 (23.6%) level one neonatal units. There were 39 (23.6%) neonatal units classified as high volume, 30 (78.0%) of which were designated level three units; consequently, 14 of the 44 (31.8%) level three designated units are not classified as high volume. Among the 20,554 infants, 1,892 (9.2%) were born in hospitals with neonatal units that are classified as high volume but not tertiary level and 1,817 (8.8%) were born in hospitals with neonatal units classified as tertiary level but not high volume.

4.5.2 Standard Logistic Regression

Table 4.2 presents the estimated odds ratios associated with admission to either tertiary or high volume neonatal care at the hospital of birth from the 'standard' logistic regression. Equivalent APEs are provided in Table 4.3. There is no statistically significant difference in the odds of mortality for very preterm infants admitted to tertiary level care at the hospital of birth compared to their counterparts admitted to non-tertiary level care. However, when considering only infants born at $\leq 26^{+6}$ weeks gestation, there is evidence of a reduction in the odds of neonatal mortality (OR: 0.65, p=0.012). This is equivalent to a 5.6 percentage point (pp) reduction in the risk of mortality (against a mortality rate of 34.2% for this group). There is no evidence of an effect on any in-hospital mortality.

For infants admitted to a high volume neonatal unit at the hospital of birth, a reduced odds of neonatal mortality is observed for those born at $\leq 32^{+6}$ weeks gestation and at $\leq 26^{+6}$ weeks gestation, but this is not replicated for infants born at 27^{+0}-32^{+6} weeks gestation. Those infants born at $\leq 26^{+6}$ weeks gestation are also at reduced odds of any in-hospital mortality (0.71, p=0.033; APE: -5.1pp) and increased odds of BPD (OR: 1.59, p=0.002; APE: +6pp) compared to their counterparts admitted to a non-high volume neonatal unit at the hospital of birth. There are no other statistically

Table 4.2 Odds ratios from standard logistic regression

	Tertiary neonatal unit			High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks	(2) $\leq 26^{+6}$ weeks	(3) $27^{+0}-32^{+6}$ weeks	(4) $\leq 32^{+6}$ weeks	(5) $\leq 26^{+6}$ weeks	(6) $27^{+0}-32^{+6}$ weeks
Neonatal mortality	0.77 (0.59-1.00)	0.65 * (0.46-0.91)	0.92 (0.69-1.22)	0.73 * (0.56-0.95)	0.62 ** (0.44-0.87)	0.86 (0.65-1.14)
Any in-hospital mortality	0.91 (0.72-1.15)	0.78 (0.57-1.06)	1.06 (0.83-1.36)	0.83 (0.65-1.05)	0.71 * (0.52-0.97)	0.96 (0.75-1.24)
BPD	1.23 ** (1.07-1.40)	1.50 ** (1.11-2.01)	1.17 (0.99-1.39)	1.11 (0.97-1.28)	1.59 ** (1.18-2.14)	1.02 (0.86-1.22)
Treatment for ROP	1.26 (0.91-1.75)	1.09 (0.76-1.57)	1.52 (0.91-2.55)	0.95 (0.68-1.32)	0.81 (0.56-1.17)	1.22 (0.71-2.09)
Surgery for NEC	1.05 (0.76-1.44)	0.89 (0.58-1.36)	1.17 (0.80-1.70)	1.05 (0.76-1.45)	0.94 (0.62-1.45)	1.11 (0.76-1.61)
PMA at discharge >40 weeks	1.17 (0.97-1.41)	1.09 (0.87-1.37)	1.19 (0.97-1.47)	1.13 (0.94-1.37)	1.11 (0.89-1.38)	1.11 (0.90-1.37)

[1] Values are odds ratios (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Table 4.3 Average partial effects from logistic regression

	Tertiary neonatal unit			High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks	(2) $\leq 26^{+6}$ weeks	(3) $27^{+0} - 32^{+6}$ weeks	(4) $\leq 32^{+6}$ weeks	(5) $\leq 26^{+6}$ weeks	(6) $27^{+0} - 32^{+6}$ weeks
Neonatal mortality	-0.0083* (-0.0161- -0.0005)	-0.0558 ** (-0.0962- -0.0153)	-0.0015 (-0.0064-0.0034)	-0.0102 * (-0.0180- -0.0024)	-0.0616 ** (-0.1020- -0.0215)	-0.0026 (-0.0075-0.0022)
Any in-hospital mortality	-0.0037 (-0.0125-0.0052)	-0.0377 (-0.0826-0.0071)	0.0013 (-0.0043-0.0069)	-0.0073 (-0.0163-0.0016)	-0.0509 * (-0.0955- -0.0064)	-0.0008 (-0.0064-0.0048)
BPD	0.0247 ** (0.0084-0.0409)	0.0578 ** (0.0178-0.0978)	0.0175 (-0.0017-0.0367)	0.0131 (-0.0036-0.0298)	0.0660 ** (0.0263-0.1060)	0.0027 (-0.0170-0.0224)
Treatment for ROP	0.0034 (-0.0015-0.0084)	0.0073 (-0.0238-0.0385)	0.0020 (-0.0006-0.0045)	-0.0008 (-0.0056-0.0041)	-0.0178 (-0.0484-0.0127)	0.0009 (-0.0016-0.0035)
Surgery for NEC	0.0007 (-0.0037-0.0050)	-0.0057 (-0.0252-0.0139)	0.0014 (-0.0021-0.0048)	0.0007 (-0.0037-0.0050)	-0.0027 (-0.0224-0.0170)	0.0009 (-0.0025-0.0043)
PMA at discharge >40 weeks	0.0119 (-0.0024-0.0262)	0.0190 (-0.0315-0.0696)	0.0096 (-0.0019-0.0210)	0.0094 (-0.0050-0.0238)	0.0237 (-0.0251-0.0724)	0.0059 (-0.0057-0.0174)

[1] Values are average partial effects (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

significant differences observed for the morbidity outcomes. Full regression results from these models are shown in Appendix B.

4.5.3 Instrumental Variable Validity

Evidence for the validity of the IVs is provided in this section. The independence assumption requires that the location of the maternal residence is not related to unobservable infant health after conditioning on the observed covariates which, importantly, include a measure of socio-economic deprivation which may be correlated with both health and the instruments. If the independence assumption is then met then we would not expect to see differences in observed characteristics between different levels of the IV (Altonji et al., 2005). Table 4.4 shows descriptive statistics for the 20,554 very preterm infants by the designation and volume of the neonatal unit nearest to the mother's place of residence. After correcting for deprivation, there are no statistically significant differences in the observed covariates.

The standard test of the exclusion restriction is the Sargan test, or its heteoskedasticity and cluster robust variant, the Hansen J test (Godfrey, 1988). These tests cannot be conducted for the non-linear specification in this chapter; however, if the model in (4.1) is re-estimated as a linear probability model then the J-statistic can be obtained. While the linear probability may not be the best approximation to the dgp, non-rejection of the validity of the instruments does provide some reassurance. In all cases the validity of the instruments was not rejected at the 5% level.

The instruments are strongly correlated with the characteristics of the unit at the hospital of birth; 88.4% of infants whose nearest neonatal unit is designated level three are born in a hospital with a level three unit compared to only 22.5% of infants whose nearest neonatal unit is not designated level three. Furthermore, an F-test of the instruments in the first stage regression yields a p-value of < 0.001.

Finally, the instruments are required to satisfy the monotonicity assumption. It cannot be verified empirically whether there are any defiers in the sample since we cannot observe the counter-factual place of birth given different values of the instrument (see

Table 4.4 Summary statistics by levels of the instrumental variables

	Designation				Volume[a]			
	Nearest unit tertiary level	Nearest unit non-tertiary level	p-value[b]	p-value[c], controlling for deprivation	Nearest unit high volume	Nearest unit non-high volume	p-value[b]	p-value[c], controlling for deprivation
n (%)	7,167 (34.9)	13,387 (65.1)			7,357 (35.8)	13,197 (64.2)		
Gestation (weeks), mean (SD)	29.6 (2.4)	29.7 (2.3)	0.040	0.418	29.6 (2.4)	29.6 (2.3)	0.181	0.526
Birth weight (g), mean (SD)	1377.4 (429.2)	1394.2 (424.5)	0.007	0.262	1376.7 (426.7)	1394.8 (425.7)	0.004	0.111
Antenatal steroids[d]	4,703 (65.6)	8,953 (66.9)	0.069	0.584	4,749 (64.6)	8,907 (67.5)	<0.001	0.052
Deprivation score-bottom 10%[e]	1,751 (24.4)	1,611 (12.0)	<0.001	NA	1,476 (20.1)	1,886 (14.3)	<0.001	NA
Male	3,820 (53.3)	7,165 (53.5)	0.761	0.854	3,958 (53.8)	7,027 (53.3)	0.447	0.378
Tertiary level unit birth[f]	4,753 (88.4)	2,290 (22.5)	<0.001	<0.001	3,839 (69.5)	3,204 (31.9)	<0.001	<0.001
High volume unit birth[f]	3,703 (68.9)	3,374 (33.1)	<0.001	<0.001	4,764 (86.3)	2,313 (23.0)	<0.001	<0.001

[1] All values are n (%) and are a proportion of the column total unless otherwise stated.
[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.
[b] Continuous variables were tested by t-test, categorical variables by chi-squared test.
[c] P-value of F-test of coefficient on instrument from a regression of variable of interest on instrument and deprivation indicator.
[d] Whether the mother received a full or partial course of antenatal steroids.
[e] Whether the mother's residence was in one of the 10% most deprived LSOAs measured by the Index of Multiple Deprivation.
[f] Birth in a hospital with either a high volume or tertiary level neonatal unit.

footnote on page 80). There is no reason to suspect the existence of defiers with the instrument, i.e. those who would deliberately impose upon themselves the burden of going to a hospital further away to go to a different volume or designation hospital than the nearest one. Certainly, there may be a small number of people who have a loyalty to a certain hospital that they would choose if a new one were built nearer, but these individuals would be never-takers. Within the sample, only 4.80% of infants were admitted to a low volume neonatal unit at the hospital of birth when the neonatal unit at the nearest hospital was a high volume unit, and only 3.97% were admitted to a low volume neonatal non-tertiary designation hospital unit at the hospital of birth when the neonatal unit at the nearest hospital was designated as tertiary level. The equivalent figures for infants born $\leq 26^{+6}$ weeks gestation are 2.85% and 1.52% respectively. These infants, who are apparent defiers, may be a random subset of infants who are not admitted to the nearest neonatal unit to the maternal residence because the nearest unit is at or near full occupancy. As such, it is unlikely that the instruments violate the monotonicity assumption. Even if this assumption was violated, given the negligible size of the possible defier population, the effects of the defiers are unlikely to have a large impact on the results. Moreover, the presence of defiers would likely lead us to underestimate the effects of high volume or tertiary designation neonatal unit at the place of birth.

4.5.4 Instrumental Variable Logistic Regression

Table 4.5 shows the estimated odds ratios using the instrumental variables logistic regressions, the equivalent APEs are shown in Table 4.6. There is no evidence for a difference in neonatal mortality between infants admitted to either tertiary or non-tertiary neonatal care at the hospital of birth. Although, there is evidence of increased odds of treatment for ROP for very preterm infants born at $27^{+0} - 32^{+6}$ weeks gestation born in a hospital with a tertiary level unit (OR: 2.17, p=0.035; APE= +0.4pp).

In contrast to the effect of tertiary level care, very preterm infants admitted to a high volume neonatal unit at the hospital of birth had significantly reduced odds of

Table 4.5 Odds ratios from instrumental variable logisitic regression

	Tertiary neonatal unit			High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks	(2) $\leq 26^{+6}$ weeks	(3) $27^{+0} - 32^{+6}$ weeks	(4) $\leq 32^{+6}$ weeks	(5) $\leq 26^{+6}$ weeks	(6) $27^{+0} - 32^{+6}$ weeks
Neonatal Mortality	0.87	1.01	0.82	0.70*	0.54**	0.80
	(0.66-1.15)	(0.63-1.61)	(0.58-1.14)	(0.53-0.92)	(0.33-0.87)	(0.56-1.13)
Any in hospital mortality	0.85	0.95	0.84	0.68**	0.51**	0.80
	(0.68-1.06)	(0.62-1.44)	(0.64-1.10)	(0.54-0.85)	(0.33-0.79)	(0.60-1.07)
BPD	1.19	1.04	1.17	1.05	1.78**	0.96
	(0.95-1.49)	(0.66-1.64)	(0.91-1.51)	(0.85-1.29)	(1.12-2.81)	(0.75-1.22)
Treatment for ROP	1.91*	1.57	2.17*	1.02	0.58	1.84
	(1.16-3.14)	(0.83-2.96)	(1.06-4.47)	(0.60-1.73)	(0.29-1.15)	(0.83-4.05)
Surgery for NEC	1.17	0.81	1.34	1.26	1.11	1.35
	(0.72-1.90)	(0.40-1.66)	(0.76-2.38)	(0.76-2.07)	(0.54-2.28)	(0.75-2.43)
PMA at discharge $>40^{+0}$ weeks	0.95	0.83	0.97	0.92	1.04	0.86
	(0.73-1.22)	(0.60-1.13)	(0.72-1.31)	(0.72-1.17)	(0.78-1.40)	(0.67-1.14)

[1] Values are odds ratios (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Table 4.6 Average partial effects from instrumental variable logisitic regression

	Tertiary neonatal unit			High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks	(2) $\leq 26^{+6}$ weeks	(3) $27^{+0} - 32^{+6}$ weeks	(4) $\leq 32^{+6}$ weeks	(5) $\leq 26^{+6}$ weeks	(6) $27^{+0} - 32^{+6}$ weeks
Neonatal mortality	-0.0044 (-0.0132-0.0045)	0.0009 (-0.0598-0.0616)	-0.0036 (-0.0095- 0.0023)	-0.0115 * (-0.0204- -0.0026)	-0.0801 * (-0.1450- -0.0156)	-0.0040 (-0.0102-0.0021)
Any in-hospital mortality	-0.0064 (-0.0152-0.0024)	-0.0080 (-0.0711-0.0551)	-0.0038 (0.0099-0.0022)	-0.0151** (-0.0240- -0.0061)	-0.1020 ** (-0.1690- -0.0035)	-0.0051 (-0.012-0.0015)
BPD	0.0210 (-0.0062-0.0482)	0.0055 (-0.0597-0.0706)	0.0180 (-0.0104-0.0464)	0.0057 (-0.0193-0.0308)	0.0819 * (0.0149-0.1490)	-0.0060 (-0.0361-0.0240)
Treatment for ROP	0.0095* (0.0018-0.0172)	0.0384 (-0.0159-0.0926)	0.0036* (0.0000-0.0073)	0.0003 (-0.0075-0.0081)	-0.0459 (-0.1020-0.0104)	0.0041* (0.0002-0.0079)
Surgery for NEC	0.0021 (-0.0045-0.0088)	-0.0096 (-0.0425-0.0234)	0.0026 (-0.0026-0.0078)	0.0031 (-0.0037-0.0099)	0.0049 (-0.0283-0.0380)	0.0024 (-0.0041-0.0089)
PMA at discharge >40 weeks	-0.0042 (-0.0235-0.0150)	-0.0425 (-0.1130-0.0279)	-0.0015 (-0.0176-0.0147)	-0.0064 (-0.0246-0.0119)	0.0078 (-0.0586-0.0743)	-0.0069 (-0.0255-0.0117)

[1] Values are average partial effects (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

neonatal mortality (OR: 0.70, p=0.011; APE= -1.1) and any in-hospital mortality (OR: 0.68, p=0.001; APE: -1.5pp). These effects are most acute amongst infants born at $\leq 26^{+6}$ weeks gestation (OR: 0.54, p=0.023; APE: -8.0pp). In terms of morbidity, the only significant effect was found for BPD (OR: 1.78, p=0.014; APE: +8.2pp) for infants born at $\leq 26^{+6}$ weeks gestation and admitted to high volume neonatal care at the hospital of birth. Full regression results from these models are shown in Appendix B.

Given that there is no longer evidence of an effect unit designation in the IV logisitic regression, it may be inferred that the effect observed in the standard logistic regression was due to its correlation with high volume.

4.6 Robustness

The results from the sensitivity analyses are presented in Tables 4.7-4.10. There are 1,172 (5.7%) infants with missing data for antenatal steroids; there are no missing values for the other covariates. The results remain qualitatively similar when all infants with any missing data are excluded from the analyses (Table 4.7).

To examine survival bias in the analyses of morbidity outcomes, infants who died are excluded from the analyses of morbidity outcomes. This does not reveal any evidence of differences in the odds ratios except for the odds of treatment for ROP for infants admitted to tertiary level care at the hospital of birth (OR: 1.96, p=0.013) (Table 4.8). No evidence of an effect for the outcome defined as any in-hospital mortality and/or BPD is observed either (Table 4.8).

We also examine the sensitivity of the results to the definition of volume, three alternative measures of volume are used. In these sensitivity analyses, the odds of any in-hospital mortality remain significantly lower for very preterm infants admitted to a high volume unit at the hospital of birth (Tables 4.9 and 4.10). Only eight hospitals (4.8%) meet the criteria of at least 100 VLBW infants per annum in any of the study years so that only a small proportion (6.5%) of the sample was inborn and admitted to these units. There is therefore imprecision around these results with wide confidence

intervals; amongst these infants, the odds of any in-hospital mortality was significantly lower but not statistically significant (Table 4.10).

4.7 Summary and Conclusions

This chapter shows that very preterm infants admitted to a high volume neonatal unit at the hospital of birth are at a lower risk of mortality than their couterparts born in and admitted to hospitals with low volume neonatal units. This in accordance with the related literature (Bartels et al., 2006; Chung et al., 2011, 2010; Cifuentes et al., 2002; Fellman et al., 2009; Johansson et al., 2004; Lasswell et al., 2010; Lorch et al., 2012; Phibbs et al., 2007; Rautava et al., 2007; Rogowski et al., 2004; Synnes et al., 2006). There are differences between the results of the standard logistic regressions and the IV logistic regressions, with the former found to generally underestimate the benefits of high-volume neonatal care at the place of birth. Moreover, the differences between the two methods suggest that an association between tertiary level designation and reduced risk of mortality is driven by high volume. The unit designation can be seen as a confounding variable in the volume analysis and vice versa. This was expected given the aim of MCNs to transfer high risk infants to high volume and designation units.

With regards to morbidity outcomes, treatment for ROP was the only morbidity for which a statistically significant effect was observed across analyses. We found that infants born at $27^{+0} - 32^{+6}$ weeks gestation in hospitals with tertiary level units were at increased odds of receiving treatment for ROP; however, only a very small number of these infants received treatment for ROP (86/17,995; 0.5%), suggesting the observed difference may not be clinically significant.

It is important to consider whether the effects that have been estimated here are causal effects or not. The policy under consideration that the evidence presented here would be used in consideration of is centralisation. If the effects in this chapter are indeed causal effects then they represent the counterfactual outcome that would result from having very preterm infants being born in hospitals with high volume neonatal

units. The instrumental variable methodology used in this chapter allows us to control for unobserved confounding that may prevent identification of causal effects. The instrumental variables used also passed the relevant validity tests. However, this chapter is unable to elucidate the mechanisms by which volume exerts a causal effect on infant clinical outcomes. As previously outlined, there are two competing explanations: economies of scale, such that the long run average costs of a high volume unit are lower; or learning by doing, where the labour force is more experienced. Given that when the results were delineated into infants born at $27^{+0} - 32^{+6}$ weeks gestation and $\leq 26^{+6}$ weeks gestation, a statistically significant reduction in the risk of mortality was only observed in the latter group, this arguably lends support to the learning by doing mechanism. If economies of scale were having a significant effect then we would expect to see reductions in mortality for all infants rather than just the most complex, rarer cases where experience may play a role. Further research is required to elucidate the mechanism. In any case, following the argument presented in Chapter 1, if learning by doing is an important factor in mediating the benefits of a high volume neonatal unit at the hospital of birth, then the benefits of a high volume neonatal unit at the place of birth are not replicable by increasing resourcing to smaller neonatal units. The skills required in the treatment of extremely preterm infants are likely to be non-substitutable and inimitable. An intervention that increases the proportion of very preterm infants born in hospitals with high-volume neonatal units may involve increasing the proportion of *in utero* transfers. Transfers of women prior to delivery are generally preferable because they are believed to be safer and less expensive than postnatal transfers of vulnerable infants (Mistry et al., 2009). However, a 2009 study showed that almost one-half of all *in utero* transfer requests to the London Ambulance Service were unsuccessful for non-clinical reasons (Gale et al., 2012a).

One of the aims of this chapter was to examine whether managed clinical networks would be able to replicate the benefits of a centralised system. As has been elucidated, the effect estimated by the instrumental variable method is valid for compliers with the instrument only, and thus we cannot conclude that MCNs are not functioning optimally for *all* infants. For those infants born near high volume units, MCNs are performing

optimally. Nevertheless, the 'standard' logistic regressions, while not able to provide evidence of a causal effect, do show evidence of an association between high volume place of birth and reduced risk of mortality for *all* infants in the sample. Thus, while these results do not conclusively reveal that MCNs are not able to provide the benefits of a more centralised system, they do provide strong evidence to suggest that is the case.

Accurate comparisons cannot be made with previous studies that have examined this same question given the multitude of differences between studies, in terms of statistical methods, institutional background, and definitions of volume. Nevertheless, these results agree with the past literature, providing evidence of a reduction in the risk of mortality for very preterm infants admitted to a high volume neonatal unit at the hospital of birth. This analysis has utilised an instrumental variable method to identify *ceteris parabis* effects of neonatal unit volume and designation by dealing with the unobserved confounding that may lead to bias in the 'standard' estimators. This can be compared to other studies which examine mortality or length of stay in neonatal units where the aim is to develop a predictive model, such as the studies by Manktelow et al. (2013) and Hinchliffe et al. (2013). Predictive models aim to have good external validity and high predictive power which is often assessed using a mean squared error criterion or similar. This is compared to the objectives of this chapter which was to minimise bias in the regression coefficient for either neonatal volume and designation. As a result these results from these studies are not readily comparable.

There are a number of limitations to this study. These are further discussed in Chapter 9. Firstly, under the instrumental variable method used in this chapter, it is not possible to estimate the effect of high volume or designation unit at the hospital of birth for non-compliers with the instrument. In this case, the non-compliers are those infants who would always go to a hospital with a different designation or volume neonatal unit to the nearest neonatal unit. As previously argued this is likely to be a very small group of infants and are possibly a random subset, given that there is no clinical reason why an infant would be a non-complier. Secondly, it has only been possible to observe infants who were admitted to neonatal units (since these are the

infants who are represented in the NNRD). It is not known what the effect would be for those infants who were born in hospital but died prior to admission to a neonatal unit. Previous studies have found evidence to suggest that, similarly to neonatal units, infants born on larger delivery suites are at a lower risk of mortality than their counterparts born elsewhere (Heller et al., 2002; Moster et al., 1999). High volume delivery suites are often in hospitals with high volume neonatal units, therefore, our results are likely to underestimate the benefit of birth in a hospital with a high volume neonatal unit, since there is also a benefit of the high volume delivery suite. Thirdly, it has not been possible to disentangle the effects of postnatal transfers on the risk of mortality. This is potentially important for policy regarding networked neonatal units. These results show that there is a benefit of admission to a high volume neonatal unit at the hospital of birth in the system as it currently stands. The effect of inter-unit transfers are just one of a multitude of pathways by which the observed figures are driven. It is not known whether an increased provision of inter-unit transfers would reduce the benefit of admission to a high volume neonatal unit at the hospital of birth seen here. Nonetheless, as discussed in the introduction, perhaps the most crucial period of an infant's care in determining clinical outcomes is the so called 'Golden hour', the first our after birth. In this period a number of team-orientated and task-based protocols are required to stabilise an infant (Doyle and Bradshaw, 2012). These tasks are complex and require skill on the part of the team administering them. It may therefore be in this period that the benefits of the experience obtained by staff working in high volume neonatal units becomes most apparent. This may suggest that an increased provision of inter-unit transfers may not eliminate the benefit of high volume neonatal units. However, further research is required.

The evidence provided in this chapter may support claims to centralise neonatal specialist healthcare in England, however, other important aspects of this care need to be taken into account, such as equity of access. An alternative policy that increases the proportion of very preterm births born in hospitals with high volume neonatal units, such as in-utero transfers, may be preferable.

Table 4.7 Results from sensitivity analysis. Infants with missing data excluded from analyses.

	Tertiary neonatal unit			High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks n=19,382	(2) $\leq 26^{+6}$ weeks n=2,452	(3) 27^{+0} – 32^{+6} weeks n=16,930	(4) $\leq 32^{+6}$ weeks n=19,382	(5) $\leq 26^{+6}$ weeks n=2,452	(6) 27^{+0} – 32^{+6} weeks n=16,930
Neonatal Mortality	0.88 (0.67-1.17)	1.03 (0.63-1.69)	0.82 (0.59-1.14)	0.68** (0.52-0.90)	0.51** (0.31-0.84)	0.80 (0.57-1.11)
Any in hospital mortality	0.85 (0.67-1.08)	0.95 (0.61-1.47)	0.84 (0.64-1.11)	0.67** (0.53-0.84)	0.50** (0.32-0.79)	0.79 (0.59-1.05)
BPD	1.16 (0.93-1.44)	1.01 (0.64-1.61)	1.15 (0.90-1.46)	1.03 (0.84-1.26)	1.86** (1.17-2.97)	0.94 (0.74-1.18)
Treatment for ROP	1.93* (1.16-3.21)	1.76 (0.91-3.77)	1.94 (0.93-4.06)	1.04 (0.61-1.77)	0.63 (0.32-1.27)	1.79 (0.81-3.95)
Surgery for NEC	1.04 (0.63-1.73)	0.68 (0.32-1.45)	1.24 (0.68-2.24)	1.24 (0.73-2.09)	1.02 (0.48-2.16)	1.38 (0.75-2.54)
PMA at discharge >40^{+0} weeks	0.94 (0.73-1.22)	0.84 (0.60-1.18)	0.97 (0.71-1.32)	0.93 (0.73-1.19)	1.06 (0.78-1.46)	0.88 (0.66-1.16)

[1] Values are odds ratios (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Table 4.8 Results from sensitivity analysis. Morbidity outcomes.

	Tertiary neonatal unit			High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks n=19,560	(2) $\leq 26^{+6}$ weeks n=1,987	(3) $27^{+0} - 32^{+6}$ weeks n=17,573	(4) $\leq 32^{+6}$ weeks n=19,560	(5) $\leq 26^{+6}$ weeks n=1,987	(6) $27^{+0} - 32^{+6}$ weeks n=17,573
BPD	1.15 (0.88-1.52)	1.07 (0.30-3.80)	1.16 (0.88-1.52)	0.93 (0.72-1.22)	0.88 (0.25-3.04)	0.94 (0.72-1.22)
Treatment for ROP	1.96* (1.15-3.32)	1.73 (0.87-3.45)	2.13* (1.04-4.40)	0.93 (0.53-1.65)	0.49 (0.23-1.03)	1.80 (0.81-3.99)
Surgery for NEC	1.12 (0.66-1.90)	0.80 (0.36-1.76)	1.29 (0.71-2.33)	1.11 (0.65-1.89)	0.82 (0.37-1.82)	1.29 (0.70-2.38)
PMA $>40^{+0}$ weeks	0.89 (0.67-1.19)	0.78 (0.53-1.15)	0.94 (0.69-1.28)	0.83 (0.63-1.08)	0.78 (0.53-1.13)	0.85 (0.63-1.13)
Any in-hospital mortality and/or BPD	1.13 (0.88-1.45)	NA[b]	1.13 (0.88-1.45)	0.92 (0.72-1.17)	0.83 (0.24-2.86)	0.92 (0.72-1.17)

[1] Values are odds ratios (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Table 4.9 Results from sensitivity analysis. Alternative measures of volume.

	High volume neonatal unit[a]: volume measured by intensive care days			High volume neonatal unit[a]: volume measured by number of admissions		
Outcome	(1) $\leq 32^{+6}$ weeks	(2) $\leq 26^{+6}$ weeks	(3) $27^{+0} - 32^{+6}$ weeks	(4) $\leq 32^{+6}$ weeks	(5) $\leq 26^{+6}$ weeks	(6) $27^{+0} - 32^{+6}$ weeks
Neonatal Mortality	0.73*	0.73	0.71*	0.81	0.78	0.82
	(0.56-0.96)	(0.45-1.19)	(0.52-0.98)	(0.61-1.06)	(0.49-1.24)	(0.59-1.13)
Any in hospital mortality	0.67**	0.65*	0.69*	0.75*	0.69	0.79
	(0.53-0.86)	(0.43-1.00)	(0.50-0.94)	(0.59-0.94)	(0.45-1.07)	(0.60-1.05)
BPD	0.98	1.28	0.93	1.09	1.41	1.02
	(0.79-1.23)	(0.81-2.02)	(0.72-1.19)	(0.88-1.35)	(0.91-2.17)	(0.79-1.32)
Surgery for ROP	0.96	0.55	1.50	1.27	0.71	1.19
	(0.56-1.57)	(0.28-1.06)	(0.66-3.43)	(0.76-2.13)	(0.36-1.42)	(0.88-4.14)
Surgery for NEC	1.16	1.11	1.22	1.10	0.95	1.15
	(0.73-1.86)	(0.54-2.28)	(0.69-2.17)	(0.67-1.81)	(0.48-1.89)	(0.63-2.13)
PMA $>40^{+0}$ weeks	0.81	0.87	0.78	0.86	0.88	0.83
	(0.63-1.04)	(0.65-1.17)	(0.58-1.04)	(0.67-1.10)	(0.64-1.21)	(0.62-1.10)

[1] Values are odds ratios (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as being in the top quartile of units.

Table 4.10 Results from sensitivity analysis. Alternative measures of volume.

	High volume neonatal unit[a]		
Outcome	(1) $\leq 32^{+6}$ weeks	(2) $\leq 26^{+6}$ weeks	(3) $27^{+0} - 32^{+6}$ weeks
Neonatal Mortality	0.40 (0.03-4.96)	NA[b]	0.74 (0.01-36.67)
Any in hospital mortality	0.28 (0.04-2.28)	1.18 (0.13-10.69)	0.52 (0.03-9.44)
BPD	1.95 (0.48-7.84)	0.29 (0.04-2.35)	1.10 (0.16-7.79)
Surgery for ROP	2.23 (0.17-29.70)	1.64 (0.07-40.08)	NA[b]
Surgery for NEC	4.11 (0.29-58.79)	0.23 (0.00-26.25)	NA[b]
PMA $>40^{+0}$ weeks	0.54 (0.11-2.64)	0.40 (0.06-2.50)	0.45 (0.05-3.95)

[1] Values are odds ratios (95% confidence intervals); * p<0.05; ** p<0.01; *** p<0.001

[2] BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birth weight z score, use of antenatal steroids, gender, infant year of birth and deprivation.

[a] High volume was defined as admitted over 100 very low birth weight (<1,500g) births per annum.

[b] Not enough observations with a positive outcome to estimate.

Chapter 5

The effect of neonatal health care expenditure on the risk of mortality among admissions to English neonatal units

Determining the relationship between healthcare expenditure and health outcomes is essential to inform health policy. Recent work in the United Kingdom has focussed on the effect of changes to healthcare spending on patient health outcomes in order to estimate cost-effectiveness thresholds for health technology assessment (HTA) (Claxton et al., 2013). This work and a number of other recent studies have generally demonstrated large, beneficial effects of healthcare expenditure on patient outcomes and have provided evidence that increased expenditure within formal healthcare settings may improve patient clinical outcomes measured by, for example, the risk of mortality (Almond et al., 2010; Claxton et al., 2013; Cutler et al., 2006; Luce et al., 2006; Martin et al., 2008; Stukel et al., 2005). However, there remains significant variability between the estimates of the health returns to healthcare spending, which may be due to the range of methods used and the variations in the expressions of the outputs of healthcare as well as the data sources. As an example of the observed variation, estimates of the cost per statistical life saved in neonatal care range from $550,000 (approxi-

mately £330,000) (Almond et al., 2010) to approximately £15 million (approximately $25 million) (Claxton et al., 2013).[1,2]

When considering whether a new medical technology should be adopted by health care providers, it is imperative to consider the opportunity cost of doing so; the adoption of a new medical technology would displace resources that could be used elsewhere to achieve health benefits. If the new technology does not at least produce health benefits of equivalent magnitude currently being achieved in the healthcare system at the same cost, then the efficiency of the health care system would be reduced by the adoption of the new technology. For the purposes of the work presented in this chapter, and the previously cited studies, the effect of primary interest is therefore the health outcomes that would result from a small change in expenditure on health care.

The recent studies that have examined this question have utilised data on expenditure and health outcomes aggregated at the local healthcare authority level (in particular, Claxton et al. (2013) and Martin et al. (2008) who used Primary Care Trust (PCT) data). However, both supply and demand side changes may affect both total healthcare expenditure and health outcomes. On the demand side, population healthcare needs and subsequent healthcare utilisation may rise, which would increase total expenditure. On the supply side, the costs of factor inputs to healthcare or the choices of inputs may both vary leading to changes in total expenditure. The effect of interest is the one driven by supply side changes to health care. Isolating the effects of supply side changes to health care expenditure from population need related changes on the level of total expenditure presents empirical difficulties. Identification of the health production function, particularly at the aggregate level, may only be possible under a set of potentially untenable assumptions. Analyses are confounded by the heterogeneity of patient populations and of medical technologies. What's more, given the scale

[1]In the study by Claxton et al. (2013), maternity and neonatal programmes of care are considered together owing to data limitations. The quoted figure relates to the combined programme for the financial years 2007/8 (table B10.3) and 2008/9 (table B11.3) rounded to the nearest million. The equivalent figure estimated using data from the financial year 2006/7 is £3.4 million (tableB8.20).

[2]The cost per statistical life, where used in this chapter, refers to the estimated change in healthcare expenditure required to reduce the risk of the mortality at the margin to reduce the total number of deaths by one.

and complexity of many healthcare systems,[3] the determination of the levels of total expenditure as well as the attribution of health outcomes to healthcare expenditure may also present issues for analyses.

The analysis in this chapter estimates the marginal effect of neonatal healthcare expenditure on the risk of mortality for newborn infants treated in neonatal units in England between 2009-13. I use these results to derive estimates of the cost per life saved and the cost per life year gained. Estimates of neonatal healthcare expenditure within neonatal units are derived from estimates of the costs per cot day from healthcare providers which are obtained from national reference cost data and are matched to individual level data extracted from the National Neonatal Research Database (NNRD). Following the empirical strategy utilised in the previous chapter, I exploit the fact that infants admitted to neonatal units are generally born in the hospital nearest to the mother's residence in order to identify the effect of increased inputs and hence expenditure on neonatal health care. I also explore neonatal unit survey data on the labour and capital inputs, and discuss issues of neonatal unit technical efficiency.

Small reductions in mortality among newborns can have large social welfare implications owing to the number of life years gained. As such, despite the often large costs associated with reducing mortality, the benefits are found to generally outweigh the costs (Cutler and Meara, 2000).[4] Indeed, 90% of the gains to life expectancy due to improvements in health care, as opposed to other public health measures, between 1950 and 2000 are attributable to reductions in infant mortality (Bunker, 1995, 2001; Bunker et al., 1994). Beyond gains in life years and health related quality of life, changes to health at birth can have long term effects on an individual's education and labour market outcomes (Black et al., 2007). Almond et al. (2010) also study the marginal returns to medical spending in this patient population but focus on only a small subset of infants—those born around the very low birth weight (VLBW; <1,500g) threshold.

[3]For the financial year 2011/12, the National Health Service in England alone employed 1.4 million people with a budget of £95.6 billion across 151 local healthcare authorities.

[4]Cutler and Meara (2000) take an estimated value per life year of $100,000 (in 2000 USD) and examine the change in expenditure on newborn infants between 1960 and 1990 along with the change in outcomes. Assuming a 3% discount rate, they estimate that between 1960 and 1990, the return to medical spending on newborn care was around 500%.

While this group of high-risk infants are of interest given the high costs they incur, they represent less than 10% of this patient population.[5]

5.1 Background

5.1.1 The National Health Service in England

Formal healthcare in England is predominantly provided by the National Health Service (NHS) which comprises a complex structure of various agencies involved in the commissioning, provision, and regulation of healthcare services. During the period of data collection for the work outlined in this chapter, April 2009 to April 2013, the structure of both the NHS in general and the specific organisation of neonatal specialist services remained relatively unchanged.[6] The structure of the NHS during this period was described in detail in Section 2.4 in Chapter 2. In this section, I briefly re-describe the structure of the NHS to clarify the nature and origin of the data used here and to inform the empirical specification that follows.

As a result of legislation in the late 1990s, between April 2002 and March 2013 independent organisations called Primary Care Trusts (PCTs) were responsible for commissioning primary, secondary, and community healthcare (Talbot-Smith and Pollock, 2006). The role of PCTs was to improve the health of their local healthcare community, to plan and secure the provision of services, and to integrate health and social care—as a result, PCTs spent around 80% of the total NHS budget (Department of Health, 2013b). For the period relevant to this study, there were 152 PCTs in England.[7] Generally, secondary and tertiary healthcare services such as neonatal specialist care was provided by individual or groups of hospitals arranged into NHS Trusts or NHS Foundation Trusts with which the PCTs contracted.

Each financial year, PCTs were allocated a certain amount of money with which

[5]This figure is derived from the National Neonatal Research Database (NNRD).

[6]Following the Health and Social Care Act (2012), the organisation of the NHS has changed. In April 2013, Primary Care Trusts were abolished and replaced with Clinical Commissioning Groups (CCGs). A full summary of changes can be found in Department of Health (2012). The proceeding discussion focusses only on PCTs but applies for the most part to CCGs.

[7]Betweeen 2002 and 2005 there were 303 prior to a restructuring.

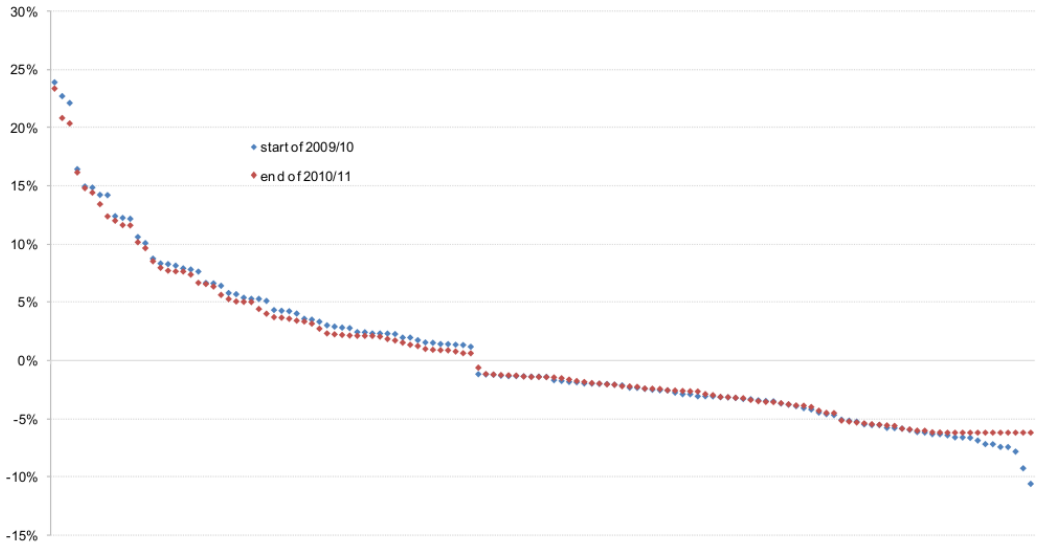
to commission healthcare services from providers. These allocations are referred to as 'target allocations' and represent the nominal amounts they should receive to meet their population's health needs; it was determined by the existing share of resources, and a combination of a funding formula and a 'pace of change' policy. The funding formula reflects a complex set of weights based upon population health needs (Department of Health, 2011). The allocation was based on the total PCT registered population weighted by a number of factors: firstly, the need for hospital and community services which was determined by age and socio-economic status among other factors; secondly, the size of the population with HIV/AIDS; thirdly, the need for prescribed drugs, determined by age, sex, and socio-economic status among other factors; and fourthly, the requirement for primary care. All of these elements were then weighted by market forces factors.

For the most part, PCT expenditure was not exactly equal to the target allocation, with some PCTs being over-target (receiving more than their allocation) while some were under-target (receiving less than their allocation). This is not unexpected given the generalisations required to determine the population healthcare needs within each PCT. Figure 5.1 shows the distances from targets in the financial years 2009/10 and 2010/11. PCTs (and now CCGs) gradually move towards their target allocations so that redistribution occurs from over-target to under-target PCTs, the rate at which this occurs is called the 'pace of change'.[8] The question then remains how or if PCTs were able to ration healthcare in order to reduce or increase expenditure.

As previously mentioned, total expenditure on healthcare within a PCT was the combination of demand side and supply side factors. On the demand side, since healthcare is provided free at the point of consumption, use cannot be regulated through a price mechanism. This is an important tenet of the NHS where ability to pay should not affect the access to healthcare. Thus, rationing is achieved through other means, including waiting lists, preventing 'non-urgent' procedures, or removing funding for low priority areas. In the short-run, the supply of healthcare is highly inelastic, further

[8]For the 2009/11 round of PCT allocations, the pace of change was set at 5.5%, implying 5.5% of the total baseline allocation was available for redistribution.

Fig. 5.1 PCT distances from target allocations



Source: House of Commons Library

emphasizing the importance of rationing through means other than the price mechanism. However, there are programmes of health care where demand cannot be 'managed' where treatment is urgent and cannot be postponed or reduced without risk of mortality, this includes emergency medicine, and in this case, neonatal medicine. For these types of healthcare excess capacity is desirable to deal with variations in demand (Madden, 1999). Nonetheless, increases in capacity have arguably been slow to keep up with increases in demand; the proportion of live births admitted to neonatal care increased from 7.6% of all live births in 2006 to 9.0% in 2011; despite this, the majority of shifts on neonatal units were found to be understaffed during this time (Pillay et al., 2011). This suggests that PCTs often had little control over demand in these cases.

5.1.2 Previous Literature

The most relevant studies related to the work presented in this chapter are the recent attempts to estimate the marginal returns to healthcare expenditure in terms of quality adjusted life years (QALYs) in order to inform the cost-effectiveness threshold employed by the National Institute for Health and Care Excellence (NICE), the health technology assessment (HTA) agency for England and Wales (Claxton et al., 2013).

This work is based on an earlier study (Martin et al., 2008). It is argued that the cost-effectiveness threshold below which new technologies are recommended for adoption should reflect the displacement implications of adoption decisions and the magnitude of the health foregone. This is estimated by determining the effect of changes to current healthcare expenditure on the health outcomes of patients at the margin. Claxton et al. (2013), as well as the preceding work by Martin et al. (2008), estimated the value of the cost-effectiveness threshold in England in terms of the cost per QALY using NHS Programme Budgeting Data. These data provide estimates of the expenditure on 23 programmes of healthcare within each of 152 local healthcare authorities. These data are described in Section 2.4 in Chapter 2. Claxton and colleagues (2013) matched relevant health outcomes to each programme of care and derived an estimate of the cost per QALY of £12,936 although there was significant uncertainty surrounding this estimate. Nevertheless, this estimate is significantly lower than the current NICE threshold, below which technologies are considered cost-effective, of £20,000 to £30,000 per QALY gained (National Institute for Health and Care Excellence, 2013). However, I argue below that that the methods used to obtain this estimate may not necessarily identify the effect of interest. Barnsley et al. (2013) also provide a critique of the assumptions used to derive the aforementioned estimate and argue that under a more realistic set of assumptions, the cost-effectiveness threshold is likely to be over £30,000 per QALY.

5.1.3 A Comment on Claxton et al. (2013)

Model Assumptions: The principal empirical estimates upon which much of the rest of the work by Claxton et al. (2013) (and of Martin et al. (2008)) are based are derived from a model of PCT expenditure decisions. The key assumptions of this model are that PCTs spend only their allocated budget and that PCTs divide expenditure between programmes of healthcare to maximise local population health. This leads to the assumption that an increase in total expenditure in one programme of healthcare must be offset with a reduction in expenditure in another. However, as figure 5.1 shows and the previous section explains, this assumption is demonstrably false. Nonetheless, a

perhaps more important issue is the interpretation of the estimated parameters, given the choice of instruments.

Instrumental Variable Interpretation: The primary effect of interest is the health forgone by the displacement of resources caused by the adoption of new health care technologies. The aim of these empirical analyses is therefore to identify the health effect of a marginal change in expenditure since this would be equivalent to the health forgone by reallocating health care resources. Consider that there are two potential mechanisms by which the causal effect of aggregate expenditure on health outcomes could be mediated. Firstly, there may be a reduction (increase) in supply side labour and capital inputs which should worsen (improve) population health outcomes; secondly, there may a change in the utilisation of health care arising through, for example, a change in the threshold for admissions or a shift in population health. In the latter case, it is due to a change in the identity of the patients being treated that expenditure and outcomes change, whereas in the former case the population being treated remains the same except the choice and level of inputs to their care changes. As another way of conceptualising the difference, consider the production possibilities frontier (PPF) of the health care system in terms of the production of population health. The adoption of a new medical technology that is less cost-effective than the threshold that is currently being achieved in the healthcare system will move the system from a point on or near the PPF to a point further away from PPF since there would be a reduction in allocative efficiency. However, a shift in population health or change in the patient cohort, leaving the mix of inputs used in health care production fixed, would lead to a shift in the curve, in whichever direction. In the case of a shift in the PPF, the effect of displacing health care resources on population health outcomes is now different, since the health outcomes that are now possible under different sets of factor inputs are now different. This implication is that shifts in expenditure identified using variables on the demand side will identify a different effect to those variables on the supply side.

Claxton et al. (2013) and Martin et al. (2008) utilise a set of census derived measures of socio-economic deprivation as instruments for expenditure in each programme

of health care.[9] One assumption is that these instruments only affect health outcomes through expenditure and have no direct effect on health outcomes otherwise—this is the exclusion assumption which is one of the conditions required for instrumental variable validity (Imbens and Angrist, 1994). To state this assumption, let h be the population health outcomes in a particular PCT, let S be the aggregate healthcare expenditure, and let z be a vector of instrumental variables. The exclusion assumption is where I implicitly assume conditioning on some set of relevant, exogenous explanatory variables:

$$h(S,z) = h(S,z') \qquad \text{for } z \neq z'$$

so that the instrumental variables do not feature directly in the health outcomes equation. Claxton et al. (2013) provide results from a test of overidentifying restrictions, which tests the null hypothesis that the instruments do not violate this assumption; the authors do not reject the null hypothesis of instrument validity. However, they note that this test may lack power under certain circumstances; as such, they additionally examine the sensitivity of their results to the case where the exclusion assumption is not met and find that, while their point estimates are not significantly altered, the uncertainty surrounding them increases to the extent where we cannot reject the null of no effect of additional healthcare expenditure.[10] Importantly though, even if the exclusion assumption is met, the mechanism by which the instruments affect total expenditure affects the interpretation, and hence validity, of the empirical results. Recall that total expenditure is determined by both supply and demand side factors. Let p be a vector of supply side determinants of S and q be an equivalent vector of demand side determinants of healthcare expenditure so that $S = S(p,q)$. Then, the health outcomes equation can be written as $h(S,z) = h(S(p,q),z)$. Since it is required that the instruments z affect S only

[9] These include, but are not limited to, the proportion of households providing unpaid care and the Index of Multiple Deprivation, a multi-dimension measure of socio-economic deprivation.

[10] These results appear in Section B7.3 of Appendix B in Claxton et al. (2013)

through p, the further restriction is required:

$$q(z) = q(z') \qquad \text{for } z \neq z'$$

where conditioning on some set of relevant, exogenous explanatory variables is assumed. This restriction is not tested. This second exclusion assumption is likely to be violated if the instruments affect the risk of developing a certain disease without affecting the clinical outcomes once this disease has been contracted. Socio-economic factors, such as those used by Claxton et al. (2013), are likely to shift population health by affecting the risk of ill health (otherwise these same factors would not be used in the determination of PCT target allocations) without necessarily affecting the clinical outcomes of patients admitted into the healthcare system. In the case of neonatal care, for example, Smith et al. (2009) showed that there was a greater burden of mortality and morbidity among infants born to mothers from more deprived areas due to increased rates of very preterm birth, but that once admitted onto neonatal units, the clinical outcomes of very preterm infants did not differ by location of the mother's residence. Socio-economic instrumental variables therefore shift the identity of the patients being treated by hospitals, which means that the effect of interest, that which operates through supply side changes to health care expenditure, may not necessarily be identified. Appendix C provides a simplified example to demonstrate this.

Attribution of Outcomes: A separate issue that it is also important to note is that the use of aggregate data by Claxton et al. (2013) may lead to issues in the attribution of health outcomes to specific programmes of care. In particular and relevantly to this paper, Claxton et al. (2013) combine maternity and neonatal programmes of care into one category and assign deaths to it on the basis of International Classification of Diseases, Version 10 (ICD-10) codes recorded on death certificates. This leads them to reallocate as many as 94% of infant deaths (deaths below one year) to other programmes of care (such as infectious diseases). Claxton et al. (2013) use this to obtain an estimate of the cost-effectiveness threshold in neonatal and maternity care of approximately £3million—two orders of magnitude larger than most other programmes

of care. However, the NNRD individual level data for 2009-13 reveal that 98% of infants to have died in neonatal units were recorded as having a neonatal ICD-10 code which may suggest mis-attribution of health outcomes by Claxton et al. (2013) for this programme of care.

In this chapter, I utilise the NNRD—a rich source of individual level data that arguably mitigates many of the concerns outlined in the preceding paragraphs.

5.1.4 Other Literature

Almond et al. (2010) estimate the incremental return to medical expenditure in a neonatal healthcare setting. Using a regression discontinuity design, exploiting a discontinuity around treatment provision to newborns either side of a 1,500g birth weight threshold, Almond et al. (2010) estimate that the cost of saving the life of a newborn with a birth weight around 1,500g is around $550,000 (in 2006 US$). Nonetheless, this result is arguably not generalisable to the wider newborn population given that only 4.1% of the admissions to neonatal specialist care are of infants with birth weights within the bandwidth utilised in the study (data from the NNRD; bandwidth 1515-1685g). In addition, as Barreca et al. (2011) discuss, there may be issues surrounding the use of a regression discontinuity design here owing to the manner in which birth weight is recorded (see Almond et al. (2011) for a reply).

Longitudinal data at the national level has been used in a more general way to obtain estimates of the effect of healthcare expenditure. Cutler et al. (2006) estimated that between 1960 and 2000 the cost per year of life gained (for the entire population) in the US was $19,900;[11] Luce et al. (2006) estimated that the return to every healthcare dollar spent in the US between 1980 and 2000 was between $1.55 and $1.94.[12] However, as Almond et al. (2010) note, while longitudinal summaries are useful, estimates of marginal returns are needed to inform policy decisions. In addition, no longitudinal

[11]This figure is unlikely to be representative of the current marginal returns to healthcare expenditure, representing, as it does, returns at a much lower level of expenditure and is therefore an infra-marginal effect.

[12]This latter result was based on the assumption of a value of a statistical life of $4 million (in 2000 US dollars). This value is similar to other studies from the US (for example, Nordhaus (2002) and other references listed in Luce et al. (2006)).

analysis exists for the UK as far as I am aware.

While the studies detailed in this section may lead us to hypothesise that there is a positive return on investment to healthcare expenditure, there are reasonable counter-arguments to suggest that there may be no effect of additional expenditure. If there are diminishing marginal returns to expenditure, there will be a point where the marginal effect of medical spending is zero. This has been termed "flat of the curve" medicine elsewhere (Fuchs, 2004). This therefore emphasizes the importance of the type of study I have undertaken, particularly given the strong pressure to reduce healthcare expenditure in the UK (Appleby, 2012).

5.2 Sample Selection

From the NNRD, data were extracted on all infants born and discharged or died between January 1st 2009 and December 31st 2013. Each infant was matched to the appropriate unit costs of the place of birth, and the nearest neonatal unit, obtained from the NHS Reference Costs. Only unit costs for IC care day (HRG4 code: 'XA01Z'), HDC care day ('XA02Z'), and SC care days ('XA03Z') were used. A full description of these datasets and the manner in which they were constructed and obtained is presented in Chapter 2.

5.3 Model and variables

I consider the following model. For each baby i, born in and admitted to the neonatal unit in hospital j in year t, let y_{ijt} be the health outcome, x_{ijt} be a vector of exogenous characteristics explaining infant health, and ex_{ijt} be the (natural logarithm of) total expenditure on neonatal health care in the neonatal unit at hospital j in year t (defined below). In addition, let α_j be a baby- and time-constant unobservable hospital effect, τ_t be year fixed effects, and let u_{ijt} be an error term. Then, I specify

$$y_{ijt} = x'_{ijt}\beta + \gamma * ex_{ijt} + \alpha_j + \tau_t + u_{ijt}. \tag{5.1}$$

The estimand of interest to this study is the average marginal effect of health care expenditure on infant health outcomes. Given that the outcome of interest at the individual level is mortality and is therefore binary, certain non-linear specifications, such as logit or probit, are preferred. However, estimators of these non-linear models are not consistent if ex_{ijt} is not independent of u_{ijt}, or if unit level unobserved heterogeneity α_j is correlated with any of the included regressors. Certain estimators have been proposed to consistently estimate models under these conditions, see for example, Papke and Wooldridge (2008). However, these generally require a balanced panel for consistency of the estimator, which I do not have in this study given the differing number of infants treated within each unit and within each year.[13] The most commonly used alternative, and the one that I shall employ for this model, is a 'fixed effect' (FE) approach which treats the unobserved heterogeneity as parameters to be estimated, although the α_j are not estimated but are eliminated as estimators of fixed effect parameters are generally inconsistent if there is not a long panel (Cameron and Trivedi, 2005a). This fixed effect approach is easily adapted to allow for the endogeneity of expenditure arising due to its possible correlation with individual and hospital unobserved heterogeneity. However, this approach is generally only possible with a linear, additive specification, which is arguably not the best approximation to the true data generating process when compared with other non-linear models. The linear model can be viewed as a non-parametric binary outcome model with linear conditional expectations, nonetheless unless all the predicted probabilities from the linear model are between zero and one, then the OLS estimator may be both biased and inconsistent (Horrace and Oaxaca, 2006). In this case, the choice of a linear model is a trade off between mis-specification and the other issues outlined here.

Total healthcare expenditure is calculated for each unit within each year. This is achieved by first totalling the number of care days provided at each level of care for each unit within each year and then multiplying by the relevant cost per cot day

[13]Woolridge (2010) proposes an extension to nonlinear correlated random effects models that allow for unbalanced panels, in particular a heteroskedastic probit for binary outcomes. In previous iterations of this research, these models were tested, however the volume of data along with other factors meant that the estimation routines for these models did not converge and were not implementable.

obtained from the Reference Cost data. The natural logarithm of total expenditure is utilised. Extraordinary procedures that are only provided by a small number of units, such as surgery, are not included, nor are costs associated with between unit transfers. While these procedures are important to the neonatal healthcare system as a whole, they do not necessarily reflect the levels of inputs to individual neonatal units and cannot be compared between units.

Within the FE framework outlined above, we are able to estimate the average effect of within unit year-to-year changes in healthcare expenditure on the risk of the health outcome of interest. Factors such as unit designation and average volume of admission are captured by the unit fixed effect. However, a correlation between within unit expenditure and infant unobserved health may arise if sicker infants are transferred to higher spending hospitals or if hospitals increase their expenditure in a particular year because they have a greater proportion of unobservably sicker infants. To deal with biases potentially arising due to this correlation, at least one instrumental variable for health care expenditure is required. This variable must satisfy the usual assumptions: it must be strongly correlated with total healthcare expenditure on the neonatal unit at the infant's hospital of birth and it must be uncorrelated with infant unobserved heterogeneity.

In this chapter, I exploit the fact that the large majority of infants are born and treated in the nearest hospital to the maternal residence. Conditional on local socioeconomic factors, the location of the maternal residence should be independent of infant unobserved health. This chapter therefore uses the total healthcare expenditure at the nearest neonatal unit to the maternal residence as an instrument for total neonatal healthcare expenditure at the neonatal unit in the hospital of birth. This aspect of the empirical strategy is the same as in the preceding chapter (published as Watson et al. (2014)). In addition, the (log) distance to the neonatal unit is also included along with its interaction with expenditure at the nearest neonatal unit, to allow for differential effects due to proximity. The validity of these instruments is tested in the usual way, an F-test of the instruments in the first stage establishes that the instruments are strongly correlated with the endogenous variable and the J-statistic from a test of

overidentifying restrictions is also reported to ensure the instruments are uncorrelated with the errors in the main equation.

5.3.1 Dependent Variable

The dependent variable used in this model is mortality which takes the value one if the infant died while admitted to a neonatal unit and zero if the infant was otherwise discharged from neonatal care.

5.3.2 Control variables

A number of exogenous determinants of in hospital mortality are included in the model, x_{ijt} in model (6.7). These are widely used in similar models (Medlock et al., 2011). The variables are gestational age and its square,[14] birth weight z-score,[15] maternal age, and dummies for whether an infant received antenatal steroids, and male sex.

In addition, I also include the local market forces factor (MFF) as a covariate. The MFF is estimated by the Department of Heath and represents the unavoidable cost differences in providing healthcare between areas, such as the cost of capital or labour inputs (Monitor, 2013). As has been emphasized throughout this chapter, the effect of interest is that due to increases or decreases to the expenditure (in real terms). However, we require this effect to be *net* of unavoidable differences in the cost of inputs since shifts in the labour market may affect the level of inputs to neonatal care without affecting the overall level of expenditure in real terms. It is for this reason the MFF is included.

5.3.3 Estimation

The linear panel instrumental variables model described in equation (6.7) can be estimated in a number of ways. Most commonly, models of this type are estimated either using the two stage least squares (2SLS) estimator (this is the one step generalised

[14]Measured using ultrasound.
[15]Birth-weight normalised within gestational age week.

methods of moments (GMM) estimator) or using the two step GMM (2SGMM) estimator. Both provide consistent estimators under the same set of assumptions, however the latter is more efficient and is used here (Cameron and Trivedi, 2005b).

To ensure consistency of the 2SGMM estimator in this framework it is necessary to assume that the instruments are strongly exogenous. In particular, let z_{ijt} be the vector of instrumental variables (which in this case has dimension 1×2), and let $\tilde{u}_{ijt} = u_{ijt} - \bar{u}_{ij}$ be the mean differenced errors from equation (6.7). Then, it is assumed that $E(z'_{ijs}\tilde{u}_{ijt}) = \mathbf{0}$ for $s,t = 1,...,T$.

For the primary analysis, observations are weighted by the inverse of the probability of being born in each of the hospitals with neonatal units in this study. Since there are more likely to be observations from high volume units, the probability of observing the unit costs in high volume units is higher, the weighting is designed to counter this. The effects of interest are the returns to medical expenditure within neonatal healthcare *as a whole*. The results are tested for sensitivity to the weighting scheme utilised in Section 5.5.

5.3.4 Other Issues

Missing data

Both the NNRD and the NHS Reference Cost Data contain missing observations (Table 2.2 in Chapter 2). Not all admissions in England are observed in the NNRD. This is due to two reasons: firstly, not all English neonatal units contribute and provide permission to use their data to the NNRD (overall, records from 165/170 units are available); secondly, the location of the maternal residence may be missing from the data set. While the data are available for the large majority of infants, there will be neonatal units for which we only observe a subset of the population. For the individual model this is only a problem if the subset of missing infants are significantly different from the subset of included infants. Since I do not possess data on the missing infants this cannot be tested empirically; however, I do not believe the infants differ since the data are likely to be either missing completely at random or missing at random

(Rubin, 1976). Similarly, NHS providers that do not provide unit cost estimates to the Reference Costs are assumed to be a random subset of providers (see table 5.1 for numbers of units).

Heterogeneous Effects

The effects of interest and those estimated from the model in (6.7) are the marginal effects of neonatal healthcare expenditure averaged over the patient population, which is γ in the model. Nonetheless, for policy purposes, as well as clinical and economic interest, the effects within certain sub-populations may also be of interest.

In many studies of mortality in neonatal units, the sample of infants under investigation is often restricted in some way. This is in part due to concerns that the mortality model may not be appropriate for all infants—those factors predicting mortality for one subset of the patient population are not successful predictors for another. Similarly, the causes of death may vary between different patient groups, which may be affected to a lesser or greater extent by increasing factor inputs to each unit. It is assumed in this chapter that increased HRG unit costs result from increased labour and capital inputs to neonatal units (this is further examined in Section 5.6). However, there are additional factors that may influence patient clinical outcomes between units above and beyond the levels of factor inputs. Infants admitted to higher volume neonatal units at the hospital of birth have been shown to have a reduced risk of mortality (Chapter 4 and Watson et al. (2014)); this may be driven, in part, by the experience of clinicians within these units. Moreover, the accumulation of specific human capital on these units may enable them to more effectively deploy the resources available in the production of neonatal healthcare, as such there may further be differences in technical efficiency between units.

The subgroups most often studied are those infants that are very low birth weight (VLBW; <1,500g) or very preterm (born at less than 33 weeks gestation) (see Chapter 3 for a review). A likelihood ratio test comparing the model in equation (6.7) estimated for the whole sample and separately for infants born at <33 weeks gestation and $\geq$ 33 weeks gestation rejected the null hypothesis of no difference ($p<0.001$). To exam-

Table 5.1 Summary statistics of the sample

Variable	Financial Year				Whole Sample
	2009/10	2010/11	2011/12	2012/13	
N_j	119	88	87	40	
N_i	34,458	27,644	30,243	9,214	101,559
Birth weight (g)	2,743.3 (929.5)	2,783.4 (895.2)	2,835.2 (894.4)	2,809.6 (893.9)	2,798.7 (902.5)
Gestational age (weeks)	36.4 (3.8)	36.7 (3.7)	36.9 (3.6)	36.7 (3.6)	36.7 (3.7)
% male	44.6	43.4	44.7	43.8	44.1
Mortality (%)	1.5	2.0	1.7	1.8	1.7
Average careday cost[a] (£)	606.64 (171.23)	639.49 (148.62)	633.48 (154.66)	652.18 (163.13)	626.72 (160.40)

N_j and N_i are the number of neonatal units and the number of individual infants respectively.
Mortality is any in hospital mortality.
Birth weight, gestational age, and unit costs are mean (sd) values.
IC Unit Costs are averaged across providers and adjusted to 2012/13 GBP using the Health Services Cost Index (HSCI)

[a] These figures are averages over neonatal units rather than infants.

ine heterogeneous effects of expenditure, I re-estimate the model for different patient groups by gestational age (the whole sample, $\leq 32^{+6}$, and $\leq 26^{+6}$). In addition, effects are estimated separately for different volume neonatal units, and levels of factor inputs are examined in Section 5.6.

The analyses are conducted in R 3.0.1 and Stata version 13.

5.4 Results

5.4.1 Summary Statistics

Table 5.1 provides summary statistics for the sample included this study. Overall, 101,559 infants were included in the sample, of which 12,559 were born at $\leq 32^{+6}$ weeks$^{+\text{days}}$ gestational age, and 2,596 at $\leq 27^{+6}$ weeks$^{+\text{days}}$ gestational age. The reported costs per cot day at all levels of care are provided in Chapter 2. The mean (sd) cost per care day over all neonatal units in the sample was £626.72 (160.40).

There are clearly a much smaller group of infants in the sample for the financial year 2012/13. This is due to a smaller group of providers submitting their estimated costs in this year (see Section 2.4.2). In Section 5.5, the main results are re-estimated

without data from this year to ensure that the attrition of units in this year is not biasing the results.

5.4.2 Instrument Validity

The validity of the instruments has been established to some extent in the previous chapter, where it was shown that the characteristics of the nearest neonatal unit to the maternal residence were as good as randomly assigned. In particular, it was shown that, conditional on the infant's socio-economic status, there was no evidence that infants differed in terms of observed characteristics by the characteristics of their nearest neonatal unit. This provides some evidence to suggest that the independence assumption is met (Altonji et al., 2005). However, such a comparison is not generally possible in the framework presented for this chapter, since the 'treatment' is continuous as is the instrumental variable. Moreover, the requirement is now that the within transformed nearest neonatal unit expenditure is conditionally independent of within transformed unobserved heterogeneity. One possible method could be to test for zero partial correlation of the within transformed instrumental variable with various within transformed observed characteristics, such as gestational age. However, only in the case where all of the variables, including the conditioning variables, are multivariate normally distributed, which is not the case here, does this imply conditional independence (Baba et al., 2004). As such, the following test of overidentifying restrictions is relied upon to provide information about instrument validity in this case.

The exclusion assumption can be tested using Hansen's J statistic. The J statistic tests the null hypothesis that the instruments are uncorrelated with the errors in model (6.7), construction of the statistic requires an overidentified model, i.e. one with a greater number of instruments than endogenous variables, which is the case here. The J-statistic is a heteroskedasicity and cluster robust form of the Sargan statistic (Godfrey, 1988). The J-statistic for each model is shown with their respective models in table 5.3. In no cases was the null hypothesis of instrument validity rejected.

The assumption that the instruments have a non-zero effect on neonatal unit expen-

diture the hospital of birth are examined in two ways. Firstly, an F-test of the excluded instruments in the first stage regression provides a p-value of <0.001 in all cases, providing strong evidence for an effect of the instruments neonatal unit expenditure at the hospital of birth. Secondly, an underidentification test— this tests the null hypothesis that the model is not identified due to irrelevant instruments—rejects this null hypothesis in all cases. The p-values are reported in table 5.3.

5.4.3 Main Results

Two sets of estimates are presented in this section. Firstly, those from the model treating health care expenditure as exogenous, as shown in table 5.2; and secondly those treating expenditure as endogenous, provided in table 5.3. In the first case, as table 5.2 shows, all of the point estimates are negative, implying that a *ceteris parabis* increase in neonatal unit expenditure is associated with a reduction in the risk of mortality. The results are not statistically significant (at the 5% level) when considering the whole sample, although the estimates are significantly different from zero in the case of very preterm and extremely preterm infants. Full regression results from these models are presented in Section C.2 in Appendix C.

It is possible that infants are transferred to neonatal units that may be more appropriate for their care or that units spend more in response to an unobservably sicker patient cohort. This may mean infants at higher risk of mortality are transferred to units with greater inputs to neonatal care.[16] If this is the case then the estimates presented in table 5.2 may be biased upwards. Table 5.3 presents results, allowing for the endogeneity of neonatal unit expenditure. All the results in this table are negative and statistically significant (at the 5% level), and below the equivalent point estimates in table 5.2. These results suggest that a 10% increase in total neonatal unit expenditure leads to a 0.2 percentage point reduction in the risk of mortality among very preterm infants (on a mortality rate of 5.0%, see table 4.1, Chapter 4.

[16]Infants are transferred to higher volume neonatal units if they are at high risk of mortality (Gale et al., 2012b), these units differ in the levels of labour and capital inputs used in the provision of health care.

Table 5.2 Regression results treating expenditure as exogenous

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
expenditure	-0.0002	-0.0160^{***}	-0.0668^{***}
	(0.001)	(0.002)	(0.008)
N	$101,559$	$12,777$	$1,729$

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses.
[2] The dependent variable is in-hospital mortality. The control variables are gestational age, gestational age squared, birth weight z-score, indicators for whether a full or partial course of antenatal steroids was administered and male sex, year fixed effects, region fixed effects, deprivation score quintile dummies, and place of birth fixed effects.
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.

Table 5.3 Regression results treating expenditure as endogenous

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
expenditure	-0.00219^{**}	-0.0227^{***}	-0.0720^{***}
	(0.000744)	(0.00198)	(0.00604)
N	101559	12776	1719
J statistic	0.761	0.897	0.694
J stat. p-value	0.102	0.118	0.461

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses.
[2] The dependent variable is in-hospital mortality. The control variables are gestational age, gestational age squared, birth weight z-score, indicators for whether a full or partial course of antenatal steroids was administered and male sex, year fixed effects, region fixed effects, deprivation score quintile dummies, and place of birth fixed effects.
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.
[4] Neonatal unit expenditure at the hospital of birth is instrumented with neonatal unit expenditure at the nearest neonatal unit to the maternal residence.

Table 5.4 Regression results treating expenditure as endogenous without inverse probability weighting estimation.

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
expenditure	−0.00243	−0.0172***	−0.0461*
	(0.001)	(0.005)	(0.018)
N	101,559	12,776	1,719
J-statistic	0.666	0.775	0.805
J stat p-val.	0.314	0.608	0.302

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses.
[2] The dependent variable is in-hospital mortality. The control variables are gestational age, gestational age squared, birth weight z-score, indicators for whether a full or partial course of antenatal steroids was administered and male sex, year fixed effects, region fixed effects, deprivation score quintile dummies, and place of birth fixed effects.
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.
[4] Neonatal unit expenditure at the hospital of birth is instrumented with neonatal unit expenditure at the nearest neonatal unit to the maternal residence.
[5] The estimator used for the primary results included an inverse probability weighting scheme for the place of birth—the results in this table do not include this weighting scheme.

5.5 Robustness and Sensitivity

In this section, I provide the results of various robustness tests of the main results presented in Section 5.4.3. In the main analyses, the results are weighted by the frequency of births in each hospital in the sample so that no one neonatal unit dominates the estimates (this is inverse probability weighting estimation). The corresponding results without using weights are presented in table 5.4. The results are qualitatively similar to those results presented in table 5.3. As a further test of the robustness, infants born in and admitted to a level one neonatal unit are excluded from the sample since these units do not (nominally) provide intensive care, these results are presented in table 5.5. Again, there is little difference between these and the main results. As a final robustness check, data from 2012/13 are excluded, since there were a low number of providers supplying reference cost data in this year, and the model re-estimated, to ensure the reduction in the number of infants that are observed for this period is not leading to inconsistency in the estimators. The results are provided in table 5.6; once again these results are qualitatively similar to those in the main results.

Table 5.5 Regression results treating expenditure as endogneous excluding level one units

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
expenditure	-0.00247^{***}	-0.0234^{***}	-0.0734^{***}
	(0.001)	(0.002)	(0.006)
N	88,335	11,648	1,604
J statistic	0.769	0.961	0.692
J stat p-val.	0.0767	0.0866	0.421

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses.
[2] The dependent variable is in-hospital mortality. The control variables are gestational age, gestational age squared, birth weight z-score, indicators for whether a full or partial course of antenatal steroids was administered and male sex, year fixed effects, region fixed effects, deprivation score quintile dummies, and place of birth fixed effects.
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.
[4] Neonatal unit expenditure at the hospital of birth is instrumented with neonatal unit expenditure at the nearest neonatal unit to the maternal residence.

Table 5.6 Regression results treating expenditure as endogenous excluding data from 2012/13

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
expenditure	-0.00211^{**}	-0.0223^{***}	-0.0692^{***}
	(0.001)	(0.002)	(0.006)
N	92,345	11,771	1,584
J-statistic	0.740	0.944	0.622
J stat. p-val	0.345	0.310	0.366

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses.
[2] The dependent variable is in-hospital mortality. The control variables are gestational age, gestational age squared, birth weight z-score, indicators for whether a full or partial course of antenatal steroids was administered and male sex, year fixed effects, region fixed effects, deprivation score quintile dummies, and place of birth fixed effects.
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.
[4] Neonatal unit expenditure at the hospital of birth is instrumented with neonatal unit expenditure at the nearest neonatal unit to the maternal residence.

5.6 Exploring the Expenditure Effect

In this section, I provide extensions to the analyses thus far presented. The purpose of this is to explore the expenditure effect observed and aid with its interpretation. The total costs incurred by a neonatal unit are a function of both the reference costs and the volume of the care the unit provides. In this section I examine associations between the average care day cost and unit labour and capital inputs; the estimated cost of a cot day reflects the level of factor inputs to neonatal unit healthcare production. I use data from a cross-sectional survey of neonatal unit labour and capital inputs to explore associations among unit costs and inputs. It may also be the case that neonatal units with the same cot day costs may have a different set of factor inputs and be more efficient than another unit. To explore this issue of technical efficiency, I re-estimate the model for different neonatal units later in this section.

The Unit Profile Survey 2011 (UPS) was a survey of English neonatal units conducted in 2011. Of 171 units surveyed, 159 (93.0%) responded. The survey aimed to collect data on labour, including nursing and physician staffing both in post and establishment, and capital, such as cots and surgical facilities.[17] The comparisons made in this section show correlations between various neonatal unit labour and capital inputs and the intensive care cot day cost. A causal effect cannot be inferred from these comparisons. It is expected that increases in staffing inputs lead to a higher unit costs, rather than vice versa, given the way the unit cost is calculated (which is described in Section 2.4.2 in Chapter 2). Furthermore, since the results in Section 5.4.3 found a significant causal effect, it may be inferred that it was due to the increased levels of inputs that mortality decreased. However, without exogenous variation in the levels of these various inputs, it is not possible to say definitively that *these inputs* are having a causal effect on mortality. Nonetheless, these comparisons are useful and do shed some light on the function of neonatal units.

From the UPS, I examine a number of variables measuring labour and capital, measured in November 2011. It is assumed that the number of whole time equivalent

[17]The UPS was a follow up to two previous surveys conducted by the MRC EPICure studies in 1997 and 2006 (Hamilton et al., 2007; Tucker, 2002).

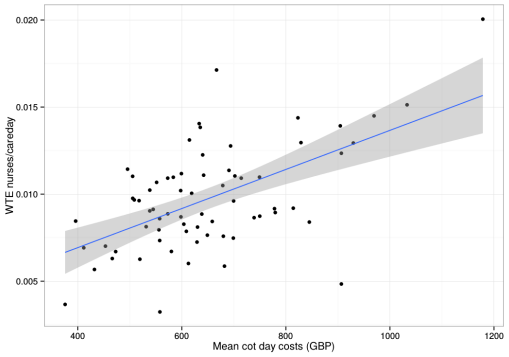
(WTE) staff and capital inputs are fixed for the year 2011/12. I explore the relationship between the number of a certain input, such as WTE nurses, per care day for the year 2011/12 and the average care day cost over the same period. For nursing, the data used are the number of whole time equivalent (WTE) nurses recorded in November 2011 which is comprised of advanced neonatal nurse practitioners (ANNPs), trained nurses, and health care assistants (HCAs) (these roles are also described in Chapter 1). These categories of nurses are disaggregated to explore the effect of the composition of nursing labour. The number of consultants is examined.[18] Capital inputs to neonatal units are wide and varied and consist of a range of durable goods, such as cots and ventilators, as well as consumables, such as drugs, canulas, and other such equipment. Consumables are not measured in the UPS, nor in the NNRD, but are generally never a limiting factor in the provision of neonatal care. To consider durable capital inputs, I use the number of neonatal cots per care day, measured at the time of the survey. These cots are further subdivided into intensive care, high dependency, and special care cots. It should be noted, though, that this measure of cots is not necessarily ideal since some of these cots may be closed, i.e. they are not utilised on the unit due to a lack of available staff to manage the cot. The ratio of closed to open cots may differ between units, however this information is not available.

The first figure, figure 6.4a, shows the correlation between the number of whole time equivalent (WTE) nurses per care day (over the financial year 2011/12) and the mean cot day cost over the same period. There is a clear positive correlation, as hypothesised; the slope coefficient is £24,517.15 and is statistically significant ($p<0.001$). This coefficient implies that an increase from the median number of nurses per care day of 0.00915 to the upper quartile of 0.01108 would increase mean cot day costs by approximately £47, or approximately 7% of the 2011 average cot day cost. If it could be assumed that the results in Section 5.4.3 could be attributed to nursing staff then this increase would lead to a 0.2 percentage point reduction in the risk of mortality among very preterm infants. Caution must be taken with these crude estimates, however, since

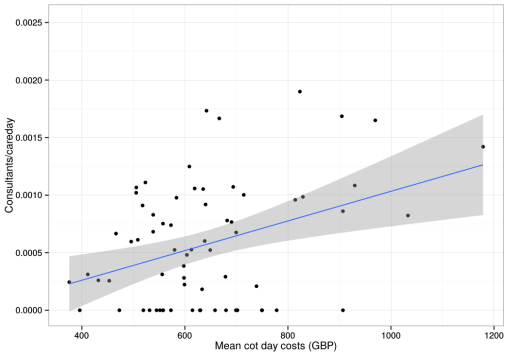
[18]Consultants in the UPS was defined as 'consultants with 50% or more of their clinical sessions (and their clinical and administrative personal assistants) dedicated to neonatal care.'

Fig. 5.2 Correaltion between average care day costs and factor inputs.

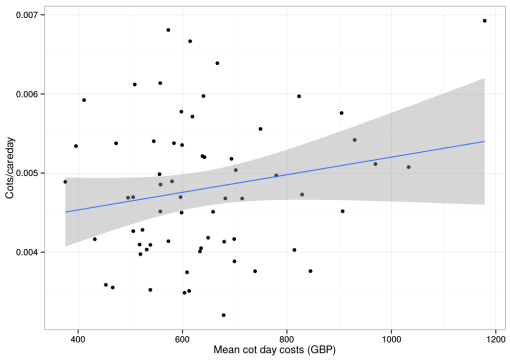
(a) WTE nurses per care day



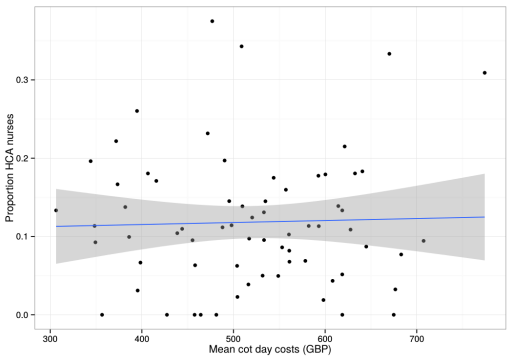
(b) Consultants oncall per care day



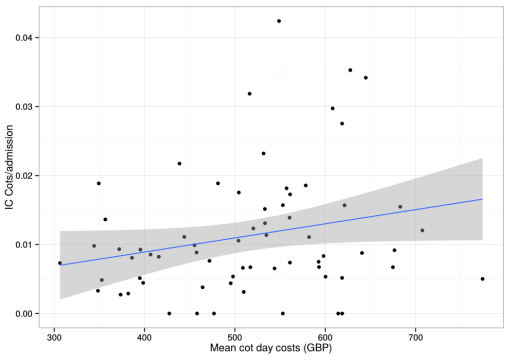
(c) Cots per care day



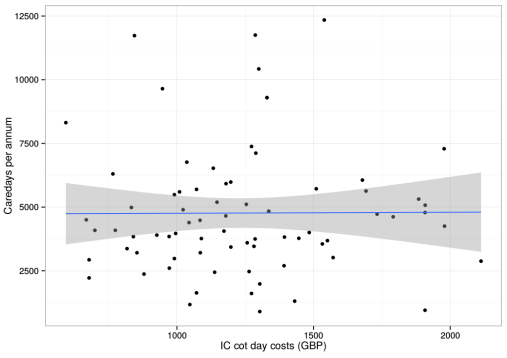
(d) Proportion of nurses that are healthcare assistants



(e) IC Cots per care day



(f) IC care days



Shaded grey areas represent 95% confidence intervals. The number of WTE staff and cots are recorded in November 2011, the number of care days considered is for the financial year 2011/12.

they rely on strong assumptions and are based on associations. A positive correlation is observed between mean cot day costs and the number of consultants per care day (figure 5.2b); but, the slope coefficient is not statistically significant (p-value 0.002). Figure 5.2c shows that a positive correlation is observed for the number of cots per admission, the measure used here to proxy the level of capital inputs per admission. The slope coefficient is not statistically significant, the p-value is 0.938.

It is possible that the composition of nursing labour also affects mean care day costs and potentially the risk of mortality. Figure 5.2d shows the correlation between mean care day costs and the proportion of the total nurse whole time equivalents (WTEs) that are HCAs—there is clearly no correlation. Figure 5.2e shows the correlation between the mean care day costs and the IC cots per care day, again there is a strong positive correlation.

As a final comparison, figure 5.2f shows the correlation between the annual number of care days provided by the neonatal unit and the intensive care unit costs. The intensive care unit costs are used here instead of the mean care day cost as the volume of a unit is strongly correlated with its composition of care days given than infants requiring intensive care are transferred to high volume neonatal units. There does not appear to be a positive correlation, the slope coefficient is not statistically significant (p=0.961).

The associations presented here only appear to find evidence of a relationship between labour inputs and the mean care day cost and not with capital inputs or volume. While it may be inferred that the effect of increased expenditure operates through varying levels of labour inputs, caution must be exercised since these are not causal effects.

Previous studies have shown very preterm infants born in hospitals with high volume neonatal units are at a lower risk of mortality than their counterparts born in hospitals with smaller neonatal units (Cifuentes et al., 2002; Phibbs et al., 2007; Watson et al., 2014). The causal effect of volume on outcomes is mediated through two mechanisms: economies of scale and learning by doing. If there were economies of scale present in the high volume neonatal units then this would imply that any given increase in the unit costs of neonatal healthcare production would correspond to a greater

Table 5.7 Regression results treating expenditure as endogenous and estimated separately for high and low volume neonatal units

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
Panel A: Results from high volume neonatal units			
expenditure	−0.0990 (0.060)	−0.499* (0.226)	−1.805* (0.755)
N	28,887	3,619	657
Panel B: Results from low volume neonatal units			
expenditure	−0.000485 (0.010)	−0.0162*** (0.003)	−0.0571*** (0.013)
N	72,672	9,157	1,060

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses.
[2] The dependent variable is in-hospital mortality. The control variables are gestational age, gestational age squared, birth weight z-score, indicators for whether a full or partial course of antenatal steroids was administered and male sex, year fixed effects, region fixed effects, deprivation score quintile dummies, and place of birth fixed effects.
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.
[4] Neonatal unit expenditure at the hospital of birth is instrumented with neonatal unit expenditure at the nearest neonatal unit to the maternal residence.

increase in factor inputs than for a smaller unit, holding external market forces constant. Alternatively, high volume units may be more technically efficient than their low volume counterparts given the same level of inputs.

To examine whether there is evidence for economies of scale or differing technical efficiency between units, I re-estimate the model in equation (6.7) separately for high volume and low volume units separately, where a 'high volume' unit is defined as a unit in the top quantile of volume by care days.[19] Table 5.7 shows the estimates of this model. The estimates from the high volume model are qualitatively larger in magnitude than those from the low volume units.

[19]This is the same definition used in the previous chapter, as well as in a previous study of neonatal units (Van Reempts et al., 2007)

Table 5.8 Estimated cost per statistical life for infants born in 2011/12

	N	Mortality (%)	Total care days	Cost (£m)	Cost per life saved (£)
Whole sample	30,243	1.7	286,921	181.764	3,005,066
$\leq 32^{+6}$	3,411	9.8	150,101	95.088	1,401,679
$\leq 27^{+6}$	685	35.9	41,114	30.480	637,171

[1] Data are taken from the sample for financial year 2011/12
[2] The average cot day cost was £633.5
[3] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.

5.7 Cost-Effectiveness Threshold

5.7.1 Cost Per Life Gained

The results in Section 5.4.3 can be easily converted into a cost per life gained.[20] I focus on the financial year 2011/12 for these calculations to prevent confounding due to between year differences in the price level and other institutional changes; this is also the year in which the UPS was conducted.

Table 5.8 shows the estimated cost per life saved based upon the results from table 5.3. Based on the total expenditure on care days provided to different groups of infants and the corresponding number of infants in that group the number of infants who would not otherwise have died with a 10% increase in expenditure is calculated. There is clearly a heterogenous effect among the different infant groups with the estimated cost per statistical life saved for very and extremely preterm infant being £1,401,679 and £637,171, respectively.

5.7.2 Cost per life year gained

To convert the above estimates into a cost per life year gained, the cost per additional life gained can be divided by the number of life years gained. I provide both discounted

[20]For the n infants in the sample, a $100/n$ percentage point change in the mortality rate is equivalent to one death. The estimated coefficients, γ, presented in the above tables provide the relationship between expenditure and the mortality rate. The change in expenditure required to realise a $100/n$ percentage point change in the mortality rate is therefore $100/\gamma n$ multiplied by the total expenditure in the sample.

and undiscounted estimates here. The key issue with this is that there are no suitable data on the life expectancies of infants admitted to neonatal units given that the survival rate of very preterm and preterm infants in the past was very low and practically negligible eighty years ago. One option is to use the average life expectancy for the English population today, which is 81 years (Office for National Statistics, 2013). However, infants born in poor health, such as those born at a low birth weight, have below average health, education, and labour market outcomes (Black et al., 2007), which may suggest a reduced life expectancy for these infants. However, the life expectancy of an infant born today is likely to be in excess of 81 years given reductions in the infant mortality rate and improvement to public health and medical care. To avoid these issues, I take the average life expectancy of 81 years. This is also the strategy of Claxton et al. (2013) in their calculations.

Using the figures presented in table 5.8, the incremental costs per life year gained for the whole sample, very preterm, and extremely preterm infants are £37,099.56, £17,304.68, and £8,384.83, respectively, not taking into account any discounting. The standard discount rate for benefits used by NICE is 3.5%. Using this rate gives equivalent costs per life year gained of £112,087.50, £52,301.46, and £23,766.17.[21] The appropriate social discount is one of the subjects of the discussion in Chapter 8.

5.7.3 Additional Costs at the Margin

The results presented in this paper suggest that any policy resulting in increased factor inputs to neonatal units, and hence unit costs, would reduce mortality among infants admitted to neonatal units. However, these infants that would have otherwise died without the policy would now generate increased expenditure owing to their requirements for care. Importantly, these infants are likely to be those that generate relatively high levels of expenditure. Going beyond care provided on a neonatal unit, after being discharged these infants will continue to generate long term costs as they require greater resources for education, healthcare, and community care than their healthy counterparts (Mangham et al., 2009). Arguably, these costs should be taken into ac-

[21] The costs are all incurred up front and so are not discounted.

count when assessing the cost per life gained in the healthcare system. Nonetheless, there is a question of whether non-healthcare costs are relevant to the cost-effectiveness of certain medicines when evaluated within a healthcare context. In any case, data on long term outcomes of infants admitted to neonatal are often not available given the low survival rates of these infants in the past, and, where they do survive, any results obtained from them are unlikely to be generalisable given the rapid progress of technology for neonatal healthcare.

In hospital costs

The additional length of stay an infant at the margin of mortality would generate is estimated in the following way. Let t be the time post-birth, and let $h(t)$ be the 'hazard' (i.e. instantaneous probability) of discharge at time t. I estimate the conditional hazard of discharge at time t using a Weibull survival model:

$$h(t|x) = \alpha t^{\alpha-1} exp(-x'\beta) \tag{5.2}$$

where $h(t|x)$ is the conditional hazard of discharge, x is the vector of exogenous variables from equation (6.7), and β and α are parameters to be estimated. This model can then be used to estimate the conditional expected length of stay for infant i with observed characteristics x_i:

$$E(t|x = x_i) = exp(-x_i'\beta/\alpha)\Gamma(\alpha^{-1} + 1). \tag{5.3}$$

where $\Gamma(.)$ is the gamma function. The estimation of the above model, by maximum likelihood, takes into account the right-censoring due to individuals dying prior to discharge (those for whom t is not observed). The expected length of stay is then estimated for each infant that died in the sample, then the difference between the predicted length of stay and the time at which the infant died is calculated. I take the median difference to be the expected increased length of stay resulting from an averted death

which in this case is 53 days.[22]

To calculate the proportion of these days that are intensive care days, model (5.2) is re-estimated for t_{IC} instead of t, where t_{IC} is the time at which intensive care provision is ceased due to a reduction in the intensity of the care provided to either HDC or SC. The median number of additional days of intensive care that an infant who died would generate is estimated at 6.7 days.

Based upon the previous estimates of the additional length of stay and the average unit costs of care for 2011: 6.7 days of intensive care cost approximately £8,070, the remaining 46.3 days, assuming they are divided equally between HDC and SC, cost approximately £31,830.[23] This gives a total of £39,990.

Post-discharge costs

In the longer term, post-discharge from a neonatal unit, preterm infants, and indeed other infants admitted to neonatal units, are at increased risk of morbidity and disability (Mangham et al., 2009; Saigal and Doyle, 2008b). This creates an additional economic burden for social and community care services as well as in the healthcare sector. Mangham et al. (2009) estimate the incremental costs associated with preterm birth (birth prior to 36 weeks gestation) of surviving to 18 years of age compared to a healthy, normal term counterpart. They find that this incremental cost is £26,752, the corresponding estimates for very preterm (<33 weeks gestation) and extremely preterm (<28 weeks gestation) births were £72,222 and £110,751.[24]

The infants on the margin of risk of mortality are likely to be at the upper end of the post-discharge costs distribution. Indeed, 66.1% of all deaths that are recorded in the sample used in this study occurred in infants who were very preterm, and 47.6% were in extremely preterm infants despite these infants making up only 10.4% and 3.2% of the patient population, respectively (data from the NNRD). I therefore assume that the incremental post-discharge costs associated with saving a life to be between £70,000

[22]This is likely to be relatively conservative since those at the margin are the most healthy of those that died and are likely to have the shortest lengths of stay among those that died.

[23]The average unit cost for high dependency care for 2011 is £868.4, and for special care it is £506.4.

[24]The figures have been inflated to 2011 GBP from the figures stated by Mangham et al. (2009) using the Health Care Services Index. This is for compatibility with the other figures quoted here.

and £110,000. Without further data it is difficult to be more precise than this.

Taking all the costs discussed in the previous section together, I obtain figures of approximately £3,140,000, £1,500,000, and £750,000 per life saved for the whole sample, very preterm, and extremely preterm infants, respectively. This translates into costs per life year, assuming an 81 year life expectancy and discounting at 3.5%, of approximately £117,120, £55,950, and £27,970 for the whole sample, very preterm, and extremely preterm infants, respectively. It is emphasized once again that these figures are relatively crude.

5.7.4 Heterogeneous Effects by Unit Volume

Table 5.7 provided estimates of the effect of neonatal healthcare expenditure at the neonatal unit at the hospital of birth disaggregated by unit volume. A has been discussed throughout this thesis, there are a two reasons to suspect that the effects of expenditure may differ between high and low volume neonatal units. In particular, high volume neonatal units may benefit from economies of scale, which would mean that an increase in expenditure would translate into greater increases in inputs to neonatal care, and high volume units may also have greater levels of specific human capital that would mean that for a given level of labour inputs the output would be greater. This latter effect of learning by doing is emphasized when considering the neonatal unit at the hospital of birth as it is the first few hours, or even the first sixty minutes—the so called Golden Hour—that may be crucial to determining the clinical outcomes of infants admitted to neonatal healthcare.

The cost per statistical life saved and the cost per life year gained, incorporating the estimates of additional costs (£140,000), are presented in table 5.9 for very preterm and extremely preterm infants. The whole sample is not included, since these estimates were not statistically significant from zero. There are clearly large differences in the estimated cost-effectiveness threshold between infants admitted to high and low volume neonatal units at the hospital of birth. The estimated cost per statistical life saved for a very preterm infants admitted to a high volume neonatal unit at the hospital of

Table 5.9 Estimated cost per statistical life for infants born in 2011/12

	High volume neonatal unit			Low volume neonatal unit		
	Cost per life	Life year gained (0% discount rate)	Life year gained (3.5% discount)	Cost per life	Cost per life year (0% discount)	Cost per life year (3.5% discount)
$\leq 32^{+6}$	201,207	2,482	7,501	1,745,050	21,543	65,089
$\leq 27^{+6}$	166,131	2,051	6,197	864,545	10,673	32,247

[1] Data are taken from the sample for financial year 2011/12
[2] The average cot day cost was £633.5
[3] High volume was defined as being in the top quartile of neonatal units by number of care days provided.
[4] Figures are rounded to the nearest thousand.
[5] $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.

birth is £201,207 compared to £1,745,050 at low volume neonatal units.

5.8 Discussion and Conclusions

This chapter has provided evidence that a *ceteris parabis* increase in neonatal unit expenditure leads to a reduction in the risk of mortality for infants born in and admitted to that unit. Increased neonatal expenditure is associated with higher nurse to patient ratios and higher cot to patient ratios. Using these results, the cost per life saved was estimated to vary for different patient groups from £637,171 for extremely preterm infants up to £3,005,066 when considering the whole cohort of admissions; however, after taking account of the additional costs that may be incurred by an infant saved whose death was averted, these estimates increased by approximately £140,000. Assuming a 3.5% discount rate, these figures translate into cost per life year gained of £27,979.64 for extremely preterm infants, £55,949.27 for very preterm infants, and £117,120.5 for all infants. I argue in Chapter 8 that a positive discount rate is inappropriate and inconsistent with some of the principles of economic evaluation; at a discount rate of zero, these values are, respectively, £37,099.56, £17,304.68, and £8,384.83.

As a weakness of this chapter, I have not been able to adjust the cost per life years gained estimate for quality to convert the results into the widely used cost per QALY gained. The life years gained by an infant who would have otherwise died on a neonatal unit are unlikely to be lived in full health. Thus, the cost per QALY gained is likely

to be higher than the cost per life year gained estimated here. This suggests that the cost per QALY in neonatal healthcare is generally above the NICE cost-effectiveness threshold of £20,000 to £30,000. It was demonstrated that the effect of expenditure differs by infant group, suggesting that increasing inputs for different patient groups may be more effective, in terms of reducing the mortality rate, rather than increasing inputs for all patients. Furthermore, the results presented in this chapter suggest that unit volume plays a role in the effectiveness with which increases to expenditure translate into patient outcomes. There were large differences observed in the effects of healthcare expenditure on the risk of mortality between infants admitted to high and low volume neonatal units at the hospital of birth. As previously discussed, unit volume may have a causal effect on patient outcomes either through economies of scale and/or increased specific human capital in the workforce. I am unable to distinguish between these two mechanisms here—this is an important topic for future research—but these results do suggest that merely increasing the expenditure in low volume neonatal units may not be a cost-effective method of improving the risk of mortality among neonatal admissions and it may not eliminate the advantage of admission to a high volume neonatal unit at the hospital of birth observed in the previous chapter. One possible interpretation of the results is that high volume units may have greater gains to realise from expenditure due to be under resourced when compared to lower volume neonatal units—the levels of inputs for a given patient may be lower in high volume neonatal units—further research is required to establish whether this is the case.

The cost per life year and cost per life year gained estimated in this chapter can be compared with those figures from the previous literature quoted in Section 5.1.2, notwithstanding the issues discussed previously. For example, Almond et al. (2011) estimated a cost per life saved among VLBW infants of $550,000 which is equivalent to approximately £330,000. This figure is of the same order of magnitude as that estimated here for high volume neonatal units at the hospital of birth, but lower than that for all neonatal units together. In addition, Almond et al. (2011) do not take into account the additional post-discharge costs associated with reducing the mortality rate. The differing values between these results and those of Almond et al. (2011) could

be explained by the composition of neonatal units in California from which Almond et al. (2011) derived their sample. There is a greater proportion of large neonatal units in California than in the UK (Phibbs et al. (2007) and Chung et al. (2010) provide data on Californian neonatal unit volume which can be compared with the results in Chapter 2). It could also be explained by the lower level of efficiency in the US health care system compared with the UK health care system as has been previously documented, differing factor input prices, or the use of alternative medical technologies. Understanding the processes by which expenditure translates into neonatal healthcare outcomes is an important future direction for this research.

As I argued in Section 5.1.3, there may be some issues with the analysis of Claxton et al. (2013), particularly with regards to neonatal care which may reduce the utility of a comparison between results. Indeed, for the combined neonatal and maternity care programme, the estimate of a cost per life saved is approximately £15million, one order of magnitude larger than that estimated here, and in Almond et al. (2011). The authors do readily admit that there are issues of data quality for some of the mortality data they use. After incorporating additional data sources and taking account of the demographics within each programme of care, the authors provide estimates for the cost per life year for the 'big four' programmes which are those programmes for which they have the highest quality data (cancer, circulatory problems, respiratory problems, and gastro-intestinal problems) as well as for all 23 programmes of care combined (these are listed in Chapter 2). These two estimates are £8,080 and £17,663 respectively (Chapter 4, table 4.9 in Claxton et al. (2013)). This chapter has demonstrated large differences in the estimates of the cost per statistical life saved and cost per life year gained when comparing different neonatal units and patient groups. This may be evidence to suggest that health technology assessment agencies should take into account the group for which different interventions are indicated and adjust the cost-effectiveness threshold appropriately.

It must be acknowledged that while the analysis presented in this chapter has a number of strengths there are also a number of weaknesses to the analysis. Many of the weaknesses are the same as those previously discussed in this thesis. In particu-

lar, there are missing data in both the NNRD sample and the reference costs data as outlined in Chapter 2. The final estimates of costs per life saved and life year gained are relatively crude calculations, owing to the lack of high quality post discharge data. In addition, the estimates here are for infants at the margin for the risk of mortality. Increased expenditure on neonatal healthcare may lead to improvements in quality of life outcomes for infants not at the margin for the risk of mortality. Not accounting for these benefits may lead me to underestimate the benefit of increased expenditure in a cost per QALY framework. Further research is required to estimate the benefits of increased expenditure to infants not at the margin for the risk of mortality. While this chapter has shown an association between different labour and capital inputs and neonatal unit expenditure, it has not shown which of the various inputs has a causal effect on mortality. Furthermore, reduction in the risk of mortality can also be achieved by increasing the technical efficiency of units, such as by changing the choice of factor inputs or how those input are employed. Further research is required to elucidate this. In the next chapter I estimate the effect of one to one nursing on the outcomes of infants who received intensive care.

Chapter 6

The effects of a one to one nurse to patient ratio on the risk of mortality in neonatal intensive care

The supply of qualified nurses has long been an important concern for healthcare policy makers (Shields, 2004). Within neonatal specialist health care, as with other areas of medicine, there are recommended nurse to patient ratios. As a particular example, and the subject of this chapter, a one to one nurse to patient ratio is recommended by the British Association of Perinatal Medicine (BAPM) for all infants receiving intensive care in Britain (British Association of Perinatal Medicine, 2010). However, much recent evidence shows that a large number of shifts in neonatal units are understaffed with respect to the recommended nurse-patient ratios; Pillay et al. (2011) observed a number of nurses over a period of five months and found that 54% of nursing shifts were understaffed with respect to the BAPM standards. This has led some groups advocating an increased nursing supply in neonatal units (Bliss, 2011). A recent systematic review of studies to examine nurse to patient ratios in a neonatal clinical care setting uncovered six studies published between 1990 and 2010. All but one of these studies found evidence that an increased nurse to patient ratio was associated with a reduced risk of adverse clinical outcomes among infants admitted to neonatal units (Sherenian et al., 2013). However, the authors of this study concluded that the studies were too

heterogeneous to support any particular nurse to patient ratio. In the previous chapter, I showed that a marginal increase to healthcare expenditure on neonatal units led to a reduction in the risk of mortality, it was also shown that expenditure on a neonatal unit was associated with increased nurse to patient ratios. The aim of this chapter is to estimate the effects of one to one nursing provision for infants receiving intensive care in neonatal units in England.

This chapter utilises the NNRD data, which provide detailed daily observations of infant care, including one to one nursing and the intensity of care provided, as well as a wide range of variables relating to the health of the infant. These data permit an individual level analysis of the effect of one to one nursing for infants that received intensive care. However, one to one nursing is highly likely to be endogenous since infants are assigned to one to one nursing on the basis on both observed and unobserved (to the analyst) health. As such, the data are aggregated and a monthly panel of neonatal units is constructed that permits estimation of the relationship between the proportion of intensive care days on which one to one nursing was provided (hereafter referred to as the one to one nursing rate) and the mortality rate.

Even within the panel data framework, the one to one nursing rate may be correlated with unobservable differences in casemix between neonatal units; however, panel data instrumental variables methods can be utilised to identify the causal effect of the one to one nursing rate (Ziliak, 1997). The method used in this chapter employs lagged one to one nursing rates as instruments for the contemporaneous one to one nursing rate. In addition, months where the neonatal unit is at a higher than average occupancy are assumed to have a smaller amount of nursing labour available per infant reducing the one to one nursing rate. The results are robust to a range of other sensitivity checks.

There is little previous literature from neonatal care to provide insight into the expected magnitude of the effect of the one to one nursing rate on the health of infants treated on neonatal units. Indeed, there is a paucity of evidence on the effect of nursing ratios in general. In addition to the six studies identified by (Sherenian et al., 2013), I only identify one further study published between 2010 and 2014 in an extension to this literature review. These studies exhibit a high degree of heterogeneity; they

use different patient populations, outcomes, and definitions of a nurse-to-patient ratio. And, while these studies generally find a positive association between lower nurse-to-patient ratios and adverse clinical outcomes, none of them can reasonably claim to have estimated a causal effect as I do here.

The rest of this chapter is structured as follows. The next section provides additional background to one to one nursing, including the relevant guidelines and literature in this area. A simple, theoretical model is presented in Section 6.2 which is used to inform the sample and variables selected in Section 6.3, and the model and hypotheses detailed in Section 6.4. The results, including sensitivity and robustness checks are shown in Section 6.5. Finally, Section 6.6 concludes.

6.1 Background

6.1.1 Clinical Guidelines

The British Association of Perinatal Medicine (BAPM) is an association that provides "services that help all those involved in perinatal practice to improve the standards of perinatal care in the British Isles." Their guidelines are the *de facto* standard for defining levels of care provided to individual infants in England. They provide recommendations regarding nurse to infant ratios that are the subject of this study. In particular, BAPM currently makes the following recommendations (the levels of care are defined in Chapter 1):

- **Intensive care**: Because of the complexities of care needed for a baby receiving intensive care, there should be 1:1 nursing.
- **High dependency care**: A nurse should not have responsibility for the care of more than two babies receiving high dependency care.
- **Special care**: A nurse should not have responsibility for more than four babies who are receiving special care.

These guidelines were developed following studies in the early 1990s that sought to determine the how nurses spent their time on a neonatal unit (Northern Neonatal Network, 1993; Williams et al., 1993). However, the role and function of a nurse within a neonatal unit is likely to have changed in the past twenty years given technological changes as well as changes to the patient population. For example, a recent study found that survival of infants born at less than 27 weeks gestation increased from 40% to 53% between 1995 and 2006 (Costeloe et al., 2012). Nonetheless, a recent study examined nursing levels within a particular MCN and found that, relative to the BAPM guidelines, 54% of shifts were understaffed (Pillay et al., 2011).

This study focusses on neonatal intensive care which has a recommended 1:1 nurse to patient ratio. As further detailed in Section 6.3, within the NNRD data utilised in this study, there is only information whether an infant received one to one nursing on a particular care day. As such, other care levels, where less intensive nursing support is usually provided, are not examined here.

6.1.2 Previous literature on Nurse to Patient Ratios in Neonatal Healthcare

The literature on nurse to patient ratios is covered in Section 3.2.1 of Chapter 3. Overall, seven studies were identified that examined nurse to patient ratios in a neonatal critical care setting (Callaghan et al., 2003; Cimiotti et al., 2006a; Filho et al., 2011; Grandi et al., 2010; Hamilton et al., 2007; Profit et al., 2010; Tucker, 2002), these are briefly discussed in further detail here along with a summary of the conclusions of Sherenian et al. (2013) who conducted a systematic review of this literature identifying six of the seven aforementioned studies.

Of the seven studies, six found that an increase in nurse to patient ratios (or conversely a decrease in patient to nurse ratios) was associated with a reduction in adverse clinical outcomes among infants admitted to neonatal units (Cimiotti et al., 2006a; Filho et al., 2011; Grandi et al., 2010; Hamilton et al., 2007; Profit et al., 2010; Tucker, 2002), while one study found the opposite effect (Callaghan et al., 2003). Two of the

studies were conducted in the United Kingdom (Hamilton et al., 2007; Tucker, 2002), two in the US (Cimiotti et al., 2006a; Profit et al., 2010), one from Australia (Callaghan et al., 2003), and two from South American nations (Filho et al., 2011; Grandi et al., 2010).

Profit et al. (2010) found that an increase of one patient per nurse led to an decrease in daily weight gain among moderately preterm infants (born at $30^{+0} - 34^{+6}$ weeks gestation) admitted to 10 neonatal intensive care units (NICUs) in California, United States. Cimiotti et al. (2006a) examined admissions to NICUs in New York and found that increases in nursing hours provided by registered nurses were associated with a reduced risk of bloodstream infections. In the United Kingdom, Hamilton et al. (2007) examined very low birth weight (born at <1,500g) or <31 weeks gestation, and found that "increasing the ratio of nurses with neonatal qualifications to intensive care and high dependency infants to 1:1 was associated with a decrease in risk-adjusted mortality of 48% (OR: 0.52, 95% CI: 0.33, 0.83)." However, these data were from 1998 to 1999, since which time the structure of neonatal care has changed in the United Kingdom. With the exception of Callaghan et al. (2003), the other three studies also found an association between increased nurse to patient ratios and a reduction in adverse clinical events. It is not clear what accounts for the opposite findings of Callaghan et al. (2003), but as Sherenian et al. (2013) note "the authors indicate that their center employs 'casual and agency staff', who are not trained in caring for high-acuity patients, at times of high census."

As Sherenian et al. (2013) discuss in their systematic review of this literature, these studies exhibited a large degree of heterogeneity. This prevented the authors from performing a pre-planned meta-analysis. Of the six studies, three examined very low birth weight (VLBW), one moderately preterm, and two all admissions; differing measures of mortality were also utilised. The measure of nurse-to-patient ratios also differed between studies; in particular, each study had a different definition of what constitutes low or high staffing to patient ratios; the high/low distinction was usually defined in terms of the sample median or other measure. Despite this, Sherenian et al. (2013) were able to conclude that "Nurse-to-patient ratios appear to affect outcomes of neonatal in-

tensive care, but limitations of the existing literature prevent clear conclusions about optimal staffing strategies."

6.2 Conceptual Model

In this section, I describe a simple conceptual model of infant health, and its relationship to one to one nursing. We consider an infant receiving intensive care in a particular neonatal unit. The neonatal unit has a fixed supply of nursing labour, L, which it allocates either to one to one nursing L_1 or to other nursing tasks L_2, so that $L_2 = L - L_1$. The health outcome of the infant, h, can be expressed as health at birth h_0 plus a scalar capturing the neonatal unit's effects on patient health, α:

$$h = h_0 + \alpha \tag{6.1}$$

The unit effect can be expressed as

$$\alpha = \gamma_0 \frac{L_2}{N} + \mathbf{1}\left[h_0 < \lambda \frac{L_1}{N}\right] \gamma_1 + a \tag{6.2}$$

which is the effect of general nursing, γ_0, plus the effect of one to one nursing, γ_1, and where N is the total need for healthcare which is the sum of total health requirements of all infants on the unit, $\mathbf{1}[.]$ is the indicator function equal to one if its argument evaluates to true and zero otherwise, a is the effect of other inputs provided by the unit, and λ is the propensity to provide one to one nursing (that may differ between units). It is assumed that there is no limit to capital inputs to healthcare, such as cots or medicines, or from other forms of labour, such as consultants, so that a is just a scalar that is not dependent on the current need for healthcare. It is assumed that infants receive one to one nursing if their health is below a threshold that is some function of the availability of nurses to provide one to one nursing. When comparing multiple units, λ may differ between units so that different units have different propensities to provide one to one nursing given the same level of available nursing labour. It is furthermore assumed that one to one nursing provides a greater benefit to an infant than other nursing tasks so

that $\gamma_1 > \gamma_0$.

I do not attempt to explicitly specify here how the allocation decisions regarding L_1 are made at the unit level. One may speculate that neonatal units are patient health maximisers and allocate labour according to the constraint $L_1 + L_2 \leq L$. Furthermore, the benefits of each type of nursing labour γ_1 and γ_2 are dependent on h_0 such that the nursing labour allocation decisions would be dependent on the overall distribution of patient health on the unit at any given time and it would need to be considered that there would be a diminishing marginal benefit of one to one nursing at the unit level. In any case, I do assume that L is fixed and that L_1 and L_2 are exogenous at the individual level and that the following derivations represent *ceteris parabis* effects.

Consider an infant at the margin of receiving one to one nursing (i.e. $h_0 = \lambda \frac{L_1}{N}$), the effect of a change in the proportion of nursing labour supply allocated to one to one nursing can be stated as follows (the effect is expressed in discrete terms owing to the discontinuity in the indicator function):

$$\frac{\Delta h}{\Delta L_1} = \frac{-\gamma_0}{N} + \frac{\gamma_1}{\Delta L_1} > 0. \tag{6.3}$$

which is composed of the beneficial effect of receiving one to one nursing and the effect due to a reduction in the available nursing labour supply to perform other tasks. From an empirical perspective, this may suggest that if the total nursing labour supply is unknown as is the case here, then the effect of one to one nursing may be underestimated.

Now, consider the level of healthcare need in the unit N (the same N as before); we can assume that if the number of infants in the unit, n, increase then the overall level of need for neonatal healthcare increases. Furthermore, if the average health of the patient population $\bar{h_0}$ declines, then the level of healthcare need increases. Therefore, we can specify

$$N = \frac{kn}{\bar{h_0}} \tag{6.4}$$

where k is a constant. Substituting this expression into equation (6.2), we can examine the effect of a change in the level of health at admission for an infant on the margin of

receiving one to one nursing:

$$\frac{\Delta h}{\Delta h_0} = 1 + \frac{\gamma_0 L_2 n}{k} - \gamma_1 \left[1 - \frac{1}{\Delta h_0} \right] \tag{6.5}$$

the sign of which is ambiguous. Similarly, for the total number of infants on the unit:

$$\frac{\Delta h}{\Delta n} = -\frac{\gamma_0 L_2 \bar{h_0}}{k n^2} - \frac{\gamma_1}{\Delta n} < 0. \tag{6.6}$$

which is negative. Supporting this, Tucker (2002) found an inverse relationship between unit occupancy and infant health. Thus, an increase in the average level of health at admission, or decrease in the number of admitted infants, increases the level of health at discharge both directly but also indirectly through reducing the total need for healthcare and therefore increasing the available labour supply conditional on a given stock of labour. In particular, more infants will receive one to one nursing.

Equations (6.3) and (6.5) make it clear that any empirical specification needs to account for the total healthcare requirement on the unit and not just one to one nursing levels. Furthermore, a number of hypotheses can be generated about the *ceteris paribis* effects of changes to certain variables, holding casemix differences fixed, that can be used to validate the empirical analyses: the effect of one to one nursing on patient health should be positive (i.e. reduce mortality), the effect of increased health at admission should be positive (i.e. reduce mortality), and an increase in the total number of admissions should be negative (i.e. increase mortality). In addition, it is expected that an increase in the volume of admissions should reduce the probability of receiving one to one nursing. This model also suggests a source of exogenous variation in one to one nursing provision—given a unit's propensity to provide one to one nursing (λ) and its available labour supply (L_1), past levels of one to one nursing may be correlated with current levels of one to one nursing.

Certainly, the model is highly simplified, with many of the assumptions arguably being unrealistic. It is unlikely that infant health and the unit effect are separable, and, it is likely the the nursing labour supply is relatively elastic with respect to changes in

the needs of the patient population as nurses may work more hours where necessary or extra staff may be hired. However, only in the case where the nursing labour supply has unit elasticity with respect to patient need, do the previous results not hold. Unit elasticity of nursing labour supply is unlikely as units are unlikely to hire enough nurses to be able to adjust perfectly to any change in patient need. This fact is emphasized by the previously cited evidence that neonatal units are understaffed in England (Pillay et al., 2011) and that there is an association between unit occupancy and patient outcomes (Tucker, 2002). It is therefore likely that nurse labour supply is relatively inelastic and so an increase in need is still expected to reduce the probability of receiving one to one nursing. Despite its simplifications, the model provides a useful foundation on which to develop an empirical specification.

6.3 Variables

6.3.1 One to one nursing

The electronic patient records from which the NNRD is derived are completed on a daily basis. However, these data were not originally collected with the explicit intention of facilitating research. Staff members input data using specialist software for each infant on a daily basis. For this reason, some variables need to be retrospectively analysed in order to determine exactly how well they were completed, and as such what they represent. The key variable of interest in the NNRD for this chapter is a binary indicator for each care day equal to one if an infant received one to one nursing and zero otherwise. An affirmative response for a particular day could represent either: whether an infant *should have* received one to one nursing, or whether an infant *did* receive one to one nursing. For the purposes of this study, we require that the one to one nursing variable represents the latter option. The purpose of this section is to determine whether this is the case. While each of the following methods of validation do not determine definitively the interpretation of the variable, together they provide strong evidence to support the claim that the variable represents whether an infant did

receive one to one nursing on a particular day. It is not possible to determine whether the variable indicates that the infant received 1:1 nursing for all 24 hours of a particular day or less.

Qualitative evidence

The uncertainty surrounding the one to one nursing variable revolves around not knowing the criteria by which staff members entering data each day choose whether to respond in the affirmative to the one to one nursing variable. As such, staff members responsible for data entry were contacted at three units by me (Unplublished Correspondence, 2013). In all units, each infant's record was completed at night by the nurse attending that infant. Staff members at all units reported that the one to one nursing variable reflects whether an infant actually received one to one nursing rather than should have received one to one nursing. In one case, the member of staff reported (where 1:1 is equivalent to one to one nursing):

> At [name of unit], it reflects whether the baby received 1:1 rather than whether it should have had 1:1. The numbers would be very different in the latter case. (Unpublished correspondence, 2013)

This comment also suggests that there should be some discrepancy between the level of one to one nursing recommended for intensive care in the guidelines and what the one to one nursing variable reflects.

Clinical Guidelines

The BAPM guidelines detailed in Section 6.1 indicate that an infant receiving intensive care should also receive one to one nursing. It is therefore possible to identify those days where an infant should have received one to one nursing which can then be compared to those days where the one to one variable is completed in the affirmative. The NNRD data show that only 10.34% of IC days were reported as 1:1 nursing days; the BAPM guidelines recommend that this figure should be 100%.

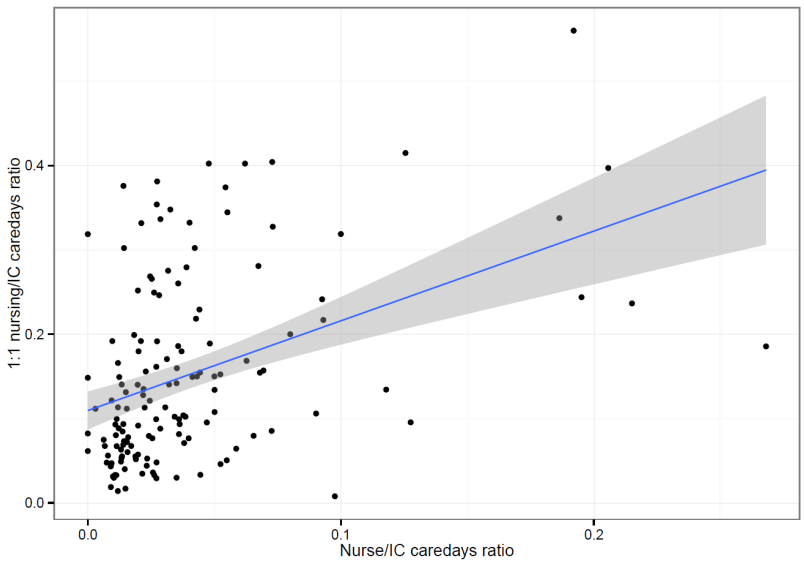
Infant severity

If the one to one nursing variable reflects whether an infant did actually receive one to one nursing rather than should have received one to one nursing then we should expect to see a relationship between the risk of mortality and the probability of receiving one to one nursing. Of all the intensive care days provided to infants who eventually died in hospital 16.97% were one to one nursing care days compared to 9.43% of the intensive care days provided to infants who survived to discharge. A test of the null hypothesis of no difference in the proportion of one to one nursing days between the two groups yielded a p-value of <0.001 providing evidence of a relationship between infant mortality and one to one nursing. In addition, 25.23% of days on which infants received surgery were one to one nursing days. Of IC days on which an infant died, 45.13% were one to one nursing days. Again, according to the BAPM guidelines, these figures should be 100%.

Nurse to patient ratios

A further test of the one to one nursing variable is to examine its relationship with average nurse to patient ratios. Arguably, the provision of one to one nursing is, at least in part, determined by the numbers of nurses employed on the unit. This obviously depends on how units prioritise nursing decisions; some may opt for lower nurse to patient ratios for healthier infants in order to provide more one to one nursing to less healthy infants whereas other units may prefer a balance in the other direction. However, it is likely that units choose a solution somewhere in between the two extremes, in which case there should be a correlation between average nurse to patient ratios and the proportion of care days that are one to one nursing days, although this correlation may not necessarily be particularly strong. In order to examine this, we use data from the Unit Profile Survey (UPS) (described in Chapter 2) which collected information on the number of whole time equivalent (WTE) medical and nursing staff within UK neonatal units for November 2011. For units completing the UPS, I extract data for IC days provided in November 2011 from the NNRD and determine the proportion of IC

Fig. 6.1 Correlation between the ratio of whole time equivalent nurses to intensive care care days and the ratio of one to one nursing days and intensive care days.



The figure includes a linear trend line (in blue) with 95% confidence interval (in grey).

days provided that were also one to one nursing care days. I also determine the ratio of WTE nurses to IC days from the UPS. Figure 6.1 shows the correlation between the two variables; the correlation coefficient is 0.42.

Transfers

As a final check, the care received by infants receiving inter-unit transfers is examined. If the 1:1 variable reflects care actually received then there should be some differences between units in the provision of 1:1 nursing owing to both different criteria used to determine whether an infant receives 1:1 nursing or not and differing labour constraints affecting their ability to actually provide 1:1 nursing. Here, I examine the proportion of infants who were entered as affirmative in the 1:1 nursing variable who received 1:1 nursing following a transfer. If the units are following guidelines, there should be very little difference either side of the transfer; whereas there should be some discrepancy if the variable reflects actual care received in the context of less than clinically optimal nursing staff inputs. Given evidence suggests that inter-unit transfers have a deleterious effect on infant health (Kelley-Quon et al., 2012; Towers et al., 2000), a null hypothesis of no variation in one to one nursing provision between units would be

equivalent to a 100% correspondence between pre- and post-transfer 1:1 nursing. The difference between pre- and post-transfer 1:1 nursing may provide some indication of the heterogeneity in 1:1 nursing provision between hospitals. The NNRD data reveal that there were 1,724 IC days between 2009 and 2011 on which an infant received both 1:1 nursing (at the unit from which an infant was transferred) and an inter-unit transfer, of these 1,254 (72.74%) received 1:1 nursing at the hospital to which the infant was transferred.

6.3.2 Outcomes

The outcome considered in this chapter is mortality. It is defined as a binary variable equal to one if the infant died in hospital and zero otherwise, for infants admitted to neonatal care. Other health outcomes were available, in particular, various illnesses associated with neonatal care. These were not utilised for this study since any analysis of these morbidity outcomes would be complicated by mortality—infants who die are censored and it is not observed whether they experience the particular morbidity of interest. Adapting the emprical methodology described in Section 6.4 to account for this requires further research.[1]

6.3.3 Control variables

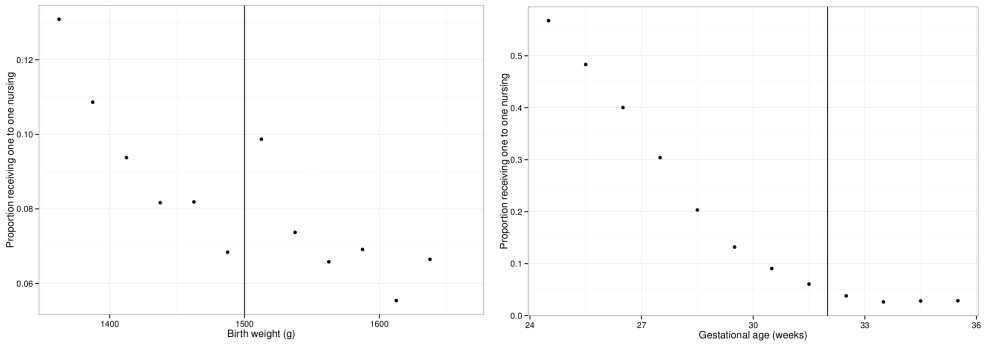
The control variables used in the analysis are the same as in the previous chapters and are gestational age and its square, birth weight z-score, maternal age, and dummies for whether an infant received antenatal steroids, and male sex.

6.4 Econometric Model and Sample

Translating the conceptual model in Section 6.2 into an empirical specification raises a number of issues. As previously discussed in Section 6.2, there are a number of factors that influence assignment to one to one nursing: infant health, unit labour sup-

[1] See also Chapter 3 for an extended discussion of methodological and empirical issues in this area.

Fig. 6.2 Relationship between infant characteristics at birth and receipt of one to one nursing



(a) Birth weight (b) Gestational age

ply, and overall level of need for care among admitted infants. In this study, we use observational data to estimate the effect of one to one nursing, using the one to one nursing indicator variable described in the previous section. However, estimation is complicated by the fact that we do not observe infant health perfectly.

There are a number of possible methods to account for the endogneity of the one to one nursing variable arising from imperfect observation of infant health. These models typically exploit some source of exogenous variation in the assignment to treatment to create a pseudo-randomisation. In the case where an instrumental variable (IV) is used for identification, this IV would be required to both affect whether an infant receives one to one nursing and be otherwise independent of unobserved infant health. Day to day labour supply and subsequent nurse-patient ratios may be candidates, however these data are not available within the NNRD or UPS. No other variable could be identified in the available datasets. Another alternative, as exploited by Almond et al. (2011) in their study of the returns to medical spending, is a discontinuity in the assignment to treatment.[2] Almond et al. (2011) showed that infants born up to 85g less than 1,500g were more likely to receive a variety of treatments than their counterparts born 85g above the 1,500g threshold. However, no such discontinuity around birth weight is evident in our data (figures 6.2a) nor is it evident when examining gestational age (Figure 6.2b).

An alternative strategy is adopted in order to estimate the effect of one to one

[2]This study is reviewed in Chapter 3.

nursing provision. The NNRD data are aggregated to the neonatal unit level for each month between January 2008 and December 2012 inclusive. This offers alternative strategies for identification. Following the model presented in Section 6.2, we estimate the following, unit level model for unit j in month t:

$$Y_{jt} = \gamma one_{jt} + x'_{jt}\beta + \pi need_{jt} + \alpha_j + \tau_t + u_{jt} \tag{6.7}$$

where Y_{jt} is the proportion of infants dying in unit j at time t, one_{jt} is the measure of one to one nursing (see Section 6.3), x_{jt} is a vector of case-mix controls to represent average health at admission (h_0 in equation (6.1) in Section 6.2), and $need_{jt}$ is the total volume of healthcare requirement in the unit (intended to capture the first term in equation (6.2)), the latter is proxied by the unit volume z-score, discussed below. The case-mix controls are taken as means of the explanatory variables of the infants assigned to unit j at time t—which infants this refers to will be discussed shortly. Additionally, α_j and τ_t are unit and year fixed effects respectively and, finally, u_{jt} is a random error term.

Estimation

Models equivalent to the panel data model specified in equation (6.7) are frequently estimated under an assumption of strict exogeneity. i.e. $E(x^*_{jt}u_{is}) = 0$, $s,t = 1,...,T$, where $x^*_{jt} = [x_{jt}, one_{jt}, need_{jt}]'$. The fixed effects or within estimator relies on this assumption, for example. However, this assumption may not hold in this case since u_{jt} contains unobserved differences in average case-mix, which are likely to be correlated with the treatment such that $E(one_{jt}u_{jt}) \neq 0$. A weaker assumption, and one that may be more likely to hold in this case, is of sequential (or weak) exogeneity, i.e. $E(x^*_{jt}u_{is}) = 0, t \leq s$. This assumption states that past values of the explanatory variables (including the treatment) are uncorrelated with the present (and future) error term, i.e. unobserved case-mix differences.

Estimation under the assumption of sequential exogeneity proceeds by first-differencing equation (6.7). After first differencing and under the sequential exogeneity assumption,

lagged values of one to one nursing can act as instruments for differenced one to one nursing. Correlation over time is assumed to arise due to a fixed unit propensity to provide one to one nursing (λ in equation (6.2)) and choice of labour allocated to one to one nursing (L_1). However, the issue here is that lagged values in the level of one to one nursing may be weak instruments for differences in one to one nursing, which may lead to inconsistent estimators. The alternative is to use the within estimator, under a strict exogeneity assumption. This estimator is biased if this assumption is not met however the bias is proportional to the reciprocal of T, which is 60 here, and so is assumed to be very small. An underidentfication test (the Kleibergen-Paap LM test) of the model estimated by first differences does not reject the null at 5% that the model is underidentified. As such, I opt for the within estimator. This is also known as fixed effects instrumental variable (FE-IV) regression.

It is also necessary to identify the number of lags of one to one nursing to use as instruments for the contemporaneous level of one to one nursing. Various methods exist for selecting the lag length for other models, such as by using an information criterion approach to assess the goodness of fit of models with varying numbers of lags. However, these have been shown to not be appropriate for panel data (Stone, 1979). Generally, choosing the appropriate number of lags is important; increasing the lag length will increase the efficiency of the estimator, whereas reducing lag length will reduce the bias in the estimator. The simplest and most widely used method is sequential testing. Beginning with the maximum number of lags, which in this case is 12 as the data are monthly, the first stage model is estimated (using the within estimator) and a fully robust t-test conducted on the last lag; if this t-statistic is not above a certain threshold (in this case 1.96) then this lag is removed and the process repeated. This method leads to four lags being used as instruments.

For the main analysis I use a two stage generalised method of moments estimator (2SGMM) since, under the assumptions here, and in the presence of intra-unit clustering, it should be both consistent and efficient (Cameron and Trivedi, 2005b). Two stage least squares (2SLS; the one step GMM estimator) estimators should also be consistent but are generally much less efficient. However, results are presented from both

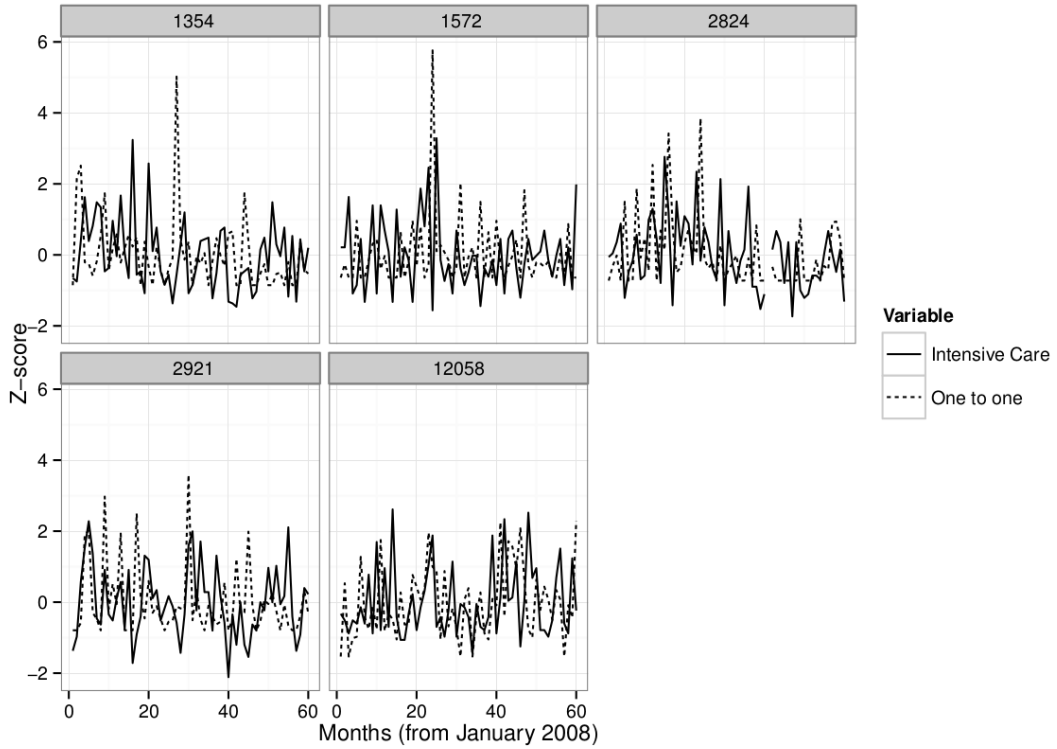
estimators—differences in the point estimates between 2SLS and 2SGMM estimators would suggest that the assumptions of the model are not met and provide evidence that both estimators are inconsistent.

The level of need for healthcare in the unit at time t features in the model (6.7); it additionally affects the probability of receiving one to one nursing. The total nurse labour supply available to a particular unit is assumed to be fixed within a particular month so that if a unit has a relatively high demand for its services in a particular month, the labour supply per patient drops and the probability of receiving one to one nursing is therefore reduced. This is reflected in Figure 6.3 which shows the standardised monthly variation in the proportion of IC care days on which one to one nursing was provided and the standardised total number of days of intensive care provided for five random units from the sample. Within each unit there appears to be a high degree of correlation between the two series. The standardised[3] total number of infants admitted for intensive care provided can act as a proxy for *need* in the model. This variable will be a valid measure of *need* and can be considered exogenous provided it is not correlated with the level of infant health. The total number of care days may not be a suitable measure of unit volume—a particularly sick infant may generate a not insubstantial proportion of a unit's care days, rendering volume as measured in care days possibly endogenous is this model.

The theoretical model in Section 6.2 allows us to make a number of testable hypotheses. Firstly, the effect of one to one nursing on mortality, γ in equation (6.7), should be negative (see equation (6.3)); secondly, the effect of *need* should be positive in equation (6.7) (see equations (6.5) and (6.6)) but have a negative effect on the probability of receiving one to one nursing (see equation (6.2); thirdly, an increase in the level of infant health should have a negative effect in equation (6.7); and, fourthly, for healthier infants who are not likely to receive one to one nursing, an increase in the proportion of one to one nursing should have a negative effect.

[3]Standardisation in this case refers to the calculation of the z-score for a variable x, $z = \frac{x-\bar{x}}{sd(x)}$.

Fig. 6.3 Variation in supply of one to one nursing and intensive care days



6.4.1 Analysis sample

There are two primary considerations to be made in selecting a sample for this analysis. In particular, which units to include and which infants within those units to include.

As discussed previously, neonatal units are classified into three groups. Level three units (neonatal intensive care units) are nominally designated to provide care to infants requiring intensive care and and according to British clinical guidelines to provide one to one nursing care. Only these units are included in the principal analysis. However, many high volume, non-level 3 units also provide a reasonable amount of intensive care since, in practice, neonatal units often do not rigidly stick to their designation for whatever reason. Figure 2.2 in Chapter 2 shows the mean monthly number of IC care days provided by level two and three units. There are clearly a number of level two units providing a volume of IC equivalent to small level three units; for example, as was detailed in Section 4.5 of Chapter 4, 9 of the 39 units classified in that chapter as high volume were designated level three (23%). As a result, I re-estimate the model using only units that are classified as high volume following the definition provided in

Chapter 4 (i.e. units in the top quartile by volume).

Only infants receiving IC are included in the sample. In the primary analysis, the unit-month explanatory variables are defined using all infants receiving intensive care within the month. The outcome is defined as the proportion of infants dying within that month; the share of infants remaining include both those who survive and those who go onto die in a following month. However, many infants will receive care in more than one month either because their care begins at the end of the month or because of a long neonatal length of stay. The data show a median length of intensive care is seven days, the upper quartile is 13 days and 12% have a duration longer than 30 days. This could lead to two problems, firstly, the errors may become serially correlated; secondly, the instruments may no longer be valid since u_{jt} may not be independent of $x_{j,t-1}$. In the first case, robust standard errors are used. In the second case, longer lags are investigated and the validity of the instruments is tested using the relevant test of overidentifying restrictions.

6.4.2 Definition of one to one nursing

The one to one nursing variable can be aggregated to the unit level in two ways. For the principal analysis in this study, the one to one care days variable is defined as the proportion of intensive care days that received one to one nursing. As an alternative definition, I also define one to one nursing as the proportion of infants receiving intensive care within a given month that received one to one nursing at any point in their care regardless of its duration.

The analyses are conducted in R 3.0.1 and Stata version 13. In particular, the package `xtivreg2` (and hence also `ivreg2`) is used in Stata.

Table 6.1 Baby level summary statistics comparing babies to have been provided one to one nursing to those that were not

Variable	One to one	None	P-value
n	20,128	49,855	
Gestational age (weeks)	32.4(5.8)	33.7(4.6)	< 0.001
Birth weight (z-score)	$-0.02(1.07)$	0.07(1.02)	< 0.001
Antenatal steroids (%)	43.7	43.5	0.698
Deprivation score, bottom 10% (%)	27.1	28.4	0.003
Male (%)	57.2	57.6	0.322
Mortality (%)	9.7	2.4	< 0.001
Level 3 place of birth (%)	51.4	56.7	< 0.001
High volume place of birth (%)	35.5	41.5	< 0.001

[a] P-values are from t-test of equality of means for continuous variables and chi-squared test for categorical variables
[1] Values are mean (sd) unless otherwise stated.

6.5 Results

6.5.1 Summary Statistics

Summary statistics at the individual level are provided in Table 6.1 and at the unit level in Table 6.2. Aggregate casemix data along with population data and numbers of admissions can be found in Table 2.1 in Chapter 2.

Table 6.1 compares the characteristics of those infants who received intensive care and any one to one nursing to those for intensive care receiving infants who did not receive any one to one nursing. Firstly, it is evident that those infants receiving one to one nursing were, on average, born at an earlier gestational age (32.4 v 33.7 weeks; p<0.001) and were born at a lower birth-weight for their gestational age (z-score -0.02 v 0.07; p<0.001). This is expected, particularly in light of the data reported in the validation exercise, in the previous section, Section 6.3. It also clear that there is a significant difference in terms of mortality with 9.4% of infants who received one to one nursing dying compared to 2.4% of those who did not receive it.

At the aggregate unit level it is possible to observe trends in the provision of one to one nursing. Table 6.2 shows that the proportion of intensive care days that are one to one nursing days has declined since 2008, from 12.4% in 2008 to 7.8% in 2012. However, the proportion of admissions onto the neonatal unit to have been provided

Table 6.2 Individual level summary statistics

Variable	2008	2009	2010	2011	2012
n	43	46	43	44	44
1:1 Caredays[a]	12.8	8.0	9.0	8.2	7.8
1:1 admissions[b]	37.7	36.4	37.5	37.2	35.9
Days per infant[c]	2.2	1.2	1.6	1.4	1.4

[a] Values are the percentage of intensive care days on which one to one nursing was provided.
[b] Values are the percentage of infants to receive any one to one nursing.
[c] Values are the mean number of one to one care days received by infants receiving intensive care.

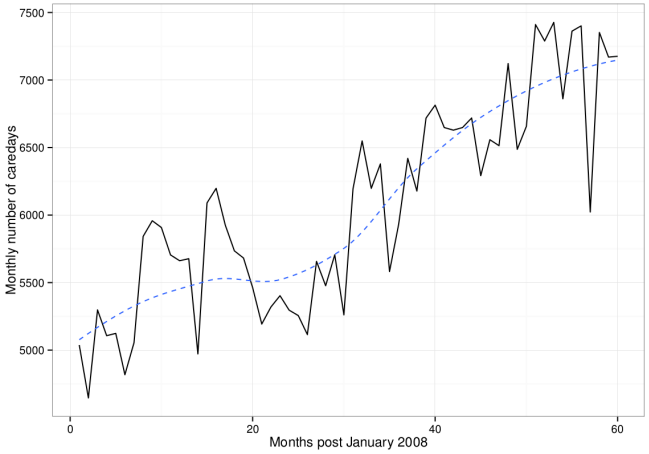
with at least one day of one to one nursing has not changed much over the same period. This is indeed reflected in by the reduction in the mean number of days of one to one nursing received by infants receiving intensive care from 2.2 days in 2008 to 1.4 days in 2012. These trends are further evidenced in figures 6.4a, 6.4b, and 6.4c, which show the change in the monthly volume of intensive care provided, the proportion of IC care days that are one to one nursing days, and the proportion of infants that received any one to one nursing.

6.5.2 IV first stage

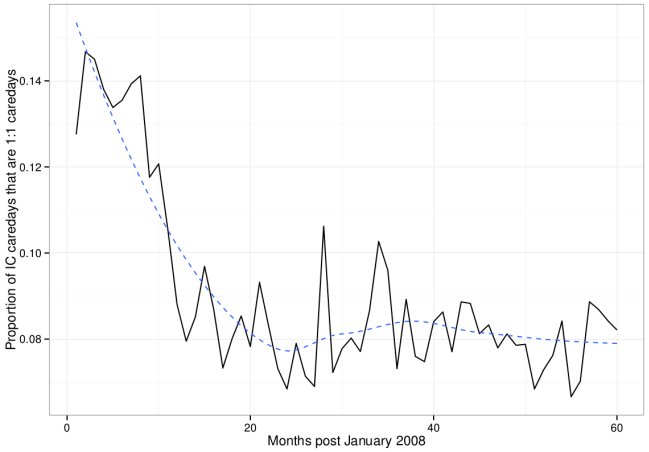
The first stage results provide information regarding both the validity of the instruments and the relationship between the instruments and the one to one nursing variables. The results from the first stage regressions are shown in Table 6.5. In the former case, an F-test of the excluded instruments in the first stage model indicates whether the excluded instruments are ‘strong’. The two F-statistics (from the model with one to one nursing measure as the proportion of intensive care days and as the proportion of admissions respectively) are 30.06 and 79.47.

The relationships between the variables in the first stage equation and the one to one nursing variable are as hypothesised. In particular, the coefficient on the intensive care volume z-score is negative and statistically significant (at the 1% level)—the intensive care volume z-score is not an instrumental variable but is shown here to demonstrate the relationship between occupancy and one to one nursing provision. A one standard deviation increase in the monthly number of infants admitted to intensive care is associated with approximately a half percentage point reduction in the proportion of

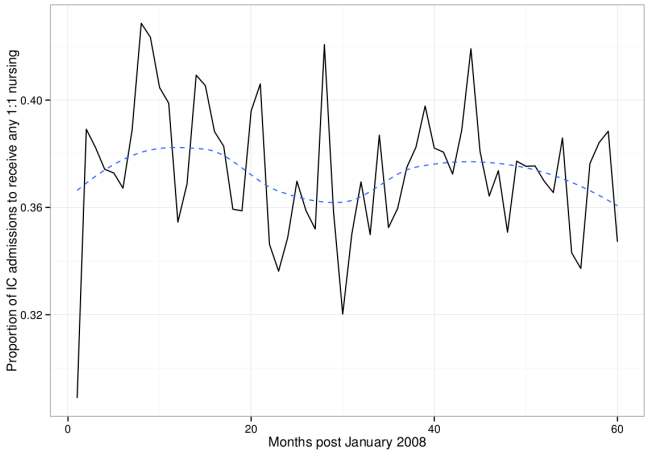
Fig. 6.4 Variation in one to one nursing over time



(a) Total intensive care days



(b) One to one nursing as a proportion of IC days



(c) Proportion of IC infants receiving any one to one

Only level three units contributing data in all years are included in the sample for these figures.

Table 6.3 Results from the first stage regression

	(1) Caredays	(2) Admissions
need	-0.00565^{***}	-0.00978^{**}
	(0.00156)	(0.00314)
one_{t-1}	0.386^{***}	0.384^{***}
	(0.0641)	(0.0252)
one_{t-2}	0.116^{*}	0.0622^{*}
	(0.0477)	(0.0240)
one_{t-3}	0.0917^{*}	0.0370
	(0.0408)	(0.0229)
one_{t-4}	0.0799^{**}	0.0452^{+}
	(0.0261)	(0.0248)
AIC	−6222.8	−3041.9
RMSE	0.0573	0.120
N	2,149	2,148

[1] + p<0.10 * p<0.05 ** p<0.01 *** p<0.001. Cluster robust standard errors in parentheses.

[2] Results from the first stage regressions. Column (1) presents results where the one to one nursing variable is defined as the proportion of intensive care days on which one to one nursing was provided. Column (2) presents results where the one to one nursing variable is defined as the proportion of admissions to intensive care to receive any one to one nursing. Regressions control for the mean values of gestational age, birth weight z-score, antenatal steroid receipt, gender, and deprivation quintile dummies. AIC = Akaike Information Criterion, RMSE = root mean squared error.

[3] *need* is not an instrumental variable in the design however it is shown here as the coefficient is of interest.

intensive care days with one to one nursing and a one percentage point reduction in the proportion of admissions with any one to one nursing.

6.5.3 Main results

Tables 6.4 and 6.5 show the estimated effect of one to one nursing on the mortality rate where one to one nursing is measured either as a proportion of total intensive care days or as a proportion of admissions, respectively. Column (1) of Table 6.4 shows the estimated effect of one to one nursing when it is treated as exogenous. It is positive and statistically significant, and implies that a ten percentage point increase in the proportion of intensive care days on which one to one nursing was provided leads to an increase in the mortality rate of 0.56 percentage points (against a mean monthly mortality rate in level three units in the sample of 5.5%). Columns (2) - (4) show the results of using instrumental variables estimators; all the estimated coefficients are negative. The point estimates in columns (2) and (3) are qualitatively similar, however the standard error in column (3) is about double that of column (2) demonstrating the efficiency gains from using the 2SGMM estimator. The main result, in column (2), suggests that a ten percentage point increase in the proportion of intensive care days receiving one to one nursing leads to a reduction in the mortality rate of 0.56 percentage points.

The results in Table 6.5 show the effects where one to one nursing is measured as the proportion of infants who receive any one to one nursing care. Again, a statistically significant reduction in the risk of mortality is observed in columns (2) and (4) (at the 10% and 5% levels respectively). Full regression results from these models are presented in Appendix D.

6.5.4 Robustness and Sensitivity

The results in the previous section are tested for robustness using a number of sensitivity analyses. Table 6.6 shows the results from these tests. The estimated coefficients in all cases are negative and qualitatively similar to those presented in the main anal-

Table 6.4 Estimated effect of one to one nursing rate on mortality rate

One to one nursing variable defined as the proportion of intensive care days on which one to one nursing was provided

	(1) OLS	(2) 2SGMM	(3) 2SLS	(4) 2SGMM
1:1 Nursing	0.0566*	−0.0562*	−0.0420	−0.0436*
	(0.0224)	(0.0281)	(0.0315)	(0.0201)
need	0.000629	0.000663	0.000370	0.000818
	(0.00116)	(0.00102)	(0.00107)	(0.000854)
RMSE	0.0574	0.0560	0.0559	0.0472
J-statistic		1.387	1.387	0.469
J-statistic p		0.709	0.709	0.926
N	2,149	2,149	2,149	1,610

[1] + p<0.10 * p<0.05 ** p<0.01 *** p<0.001. Cluster robust standard errors in parentheses.

[2] Column (1) is estimated using an OLS estimator (OLS) and treats nursing as exogenous. Columns (2) and (4) are estimated by two stage GMM (2SGMM), and column (3) by two stage least sqaures (2SLS). Columns (1)-(3) use data from level three units while column (4) uses data from high volume units. The dependent variable in each case is the mortality rate (measured between zero and one). Estimates are interpreted as the percentage point changing in the mortality rate resulting from a one percentage point increase in the proportion of intensive care days with one to one nursing. Regressions control for the mean values of gestational age, birth weight z-score, antenatal steroid receipt, gender, and deprivation quintile. RMSE = root mean squared error.

Table 6.5 Estimated effect of one to one nursing rate on mortality rate

One to one nursing variable defined as the proportion of infants receiving intensive care to have been provided any one to one nursing

	(1) OLS	(2) 2SGMM	(3) 2SLS	(4) 2SGMM
1:1 Nursing	0.0241*	−0.0299+	−0.0233	−0.0291*
	(0.0113)	(0.0160)	(0.0173)	(0.0113)
need	0.000583	0.000297	0.000334	0.000487
	(0.00115)	(0.000951)	(0.00109)	(0.000769)
RMSE	0.0574	0.0561	0.0560	0.0474
J-statistic		2.082	2.082	4.354
J-statistic p		0.556	0.556	0.226
N	2,148	2,148	2,148	1,609

[1] + p<0.10 * p<0.05 ** p<0.01 *** p<0.001. Cluster robust standard errors in parentheses.

[2] Column (1) is estimated by taking first differences and using an OLS estimator (FD-OLS) and treats nursing as exogenous. Columns (2) and (4) are estimated by two stage GMM (2SGMM), and column (3) by two stage least sqaures (2SLS). Columns (1)-(3) use data from level three units while column (4) uses data from high volume units. The dependent variable in each case is the mortality rate (measured between zero and one). Estimates are interpreted as the percentage point changing in the mortality rate resulting from a one percentage point increase in the proportion of intensive care days with one to one nursing. Regressions control for the mean values of gestational age, birth weight z-score, antenatal steroid receipt, gender, and deprivation quintile. RMSE = root mean squared error.

ysis, the estimated effect of a ten percent increase in one to one nursing provision is a reduction in the mortality rate of between 0.37 and 0.67 percentage points. However, in two of six of the models, the results are no longer statistically significant.

6.6 Discussion and Conclusions

In this chapter, I found that increases in the proportion of intensive care days that are one to one care days reduces the mortality rate in neonatal units. The results from analyses where an alternative measure of one to one nursing rate, along with the results from a variety of robustness checks were all qualitatively similar, supporting the main findings. A number of previous studies also found an association between higher nurse to patient ratios and lower rates of adverse clinical outcomes within neonatal units (Sherenian et al., 2013). Given the findings of previous studies, and the existing guidelines which stipulate a one to one nurse to patient ratio for infants receiving neonatal intensive care in the United Kingdom (British Association of Perinatal Medicine, 2010), it was expected that higher one to one nursing rates would be associated with lower mortality rates. However, the effects estimated here are marginal effects, and represent the expected impact on the mortality rate by increasing one to one nursing provision marginally. It should be considered that there is a diminishing marginal benefit of one to one nursing, such that nursing labour may be more effective if assigned to other tasks, beyond a certain point. Nonetheless, the evidence presented here along with the previously published evidence suggests that a policy of increasing one to one nursing for neonates receiving intensive care would be beneficial for those patients.

The analysis in this chapter has a number of strengths when compared to other studies on this topic. This is the first such study to take into account unobservable differences between units and as such is the first to arguably estimate causal effects. One to one nursing, in the context of this study, can be seen as a treatment provided to infants who receive intensive care, which sicker infants may be more likely to receive. The other studies to examine nurse to patient ratios on neonatal units (Callaghan et al.,

Table 6.6 Estimates from sensitivity analyses

	(1)	(2)	(3)	(4)	(5)	(6)
1:1 Nursing	−0.0373	−0.0674*	−0.0668	−0.0566*	−0.0437*	−0.0451*
	(0.0321)	(0.0307)	(0.0424)	(0.0283)	(0.0222)	(0.0204)
Instruments[a]						
Lags	1 – 3	1 – 5	2 – 4	1 – 4	1 – 4	1 – 4
Any IC z-score	*X*	*X*			*X*	*X*
Nearest IC z-score			*X*			
Sample[b]						
IC vol measure				*Caredays*	*Admissions*	
Min. IC vol					100	15
RMSE	0.0560	0.0563	0.0561	0.0561	0.0507	0.0478
Hansen J-statistic	0.477	4.528	1.310	1.378	0.542	0.979
Hansen J-statistic p	0.788	0.339	0.519	0.711	0.909	0.806
N	2,194	2,104	2,149	2,135	1,819	1,766

[1] + p<0.10 * p<0.05 ** p<0.01 *** p<0.001. Cluster robust standard errors in parentheses.

[2] The dependent variable in each case is the mortality rate (measured between zero and one), the 1:1 nursing variable is defined as the proportion of infants receiving intensive care to have received any one to one nursing (between zero and one). Estimates are interpreted as the percentage point changing in the mortality rate resulting from a one percentage point increase in the proportion of intensive care days with one to one nursing. Regressions control for the mean values of gestational age, birth weight z-score, antenatal steroid receipt, gender, and deprivation quintile. AIC = Akaike Information Criterion, RMSE = root mean squared error.

[a] 'Lags' indicates which lags of the 1:1 nursing variable are used as instruments: 2-3 indicates that $one_{j,t-2}$ and $one_{j,t-3}$ are used as instruments.

[b] Columns (5) and (6) show the results from varying the 'high volume' threshold as indicated.

2003; Cimiotti et al., 2006a; Filho et al., 2011; Grandi et al., 2010; Hamilton et al., 2007; Profit et al., 2010; Tucker, 2002), examined the effect of the nurse to patient ratio overall rather than any specific allocation of nursing labour. Overall nurse labour provision may be only weakly correlated with fluctuations in the severity of the patient casemix, mitigating concerns of endogeneity in these studies, to some extent. Taken together with the findings of this chapter, these previous studies provide evidence to support a policy of increased nurse to patient ratios. However, this research does not imply that one to one nursing is the optimal ratio for these patients, rather than one to one nursing is preferable to less than one to one for these patients.

There may be a greater concern in this study with the endogeneity of the nurse to patient ratio than with previous studies. One to one nursing, in the context of this study, can be seen as a treatment provided to infants who receive intensive care, sicker infants may be more likely to receive one to one nursing. The other studies to examine nurse to patient ratios on neonatal units (Callaghan et al., 2003; Cimiotti et al., 2006a; Filho et al., 2011; Grandi et al., 2010; Hamilton et al., 2007; Profit et al., 2010; Tucker, 2002), examined the effect of the nurse to patient ratio overall rather than any specific allocation of nursing labour. Differences between units in terms of the number of nurses per patient may arise for reasons other than casemix differences, such as the level of funding for labour allocated to the unit. This may somewhat mitigate concerns about the endogeneity of nurse to patient ratios in these studies. In any case, this analysis provides some evidence to support the current BAPM guidelines of one nurse for one patient in intensive care. However, this research does not imply that one to one nursing is the optimal ratio for these patients, rather than one to one nursing is preferable to less than one to one for these patients.

Another strength of this study is the comprehensiveness of the NNRD data utilised. The two previous studies that were conducted in the United Kingdom, Tucker (2002) and Hamilton et al. (2007), also utilised comprehensive data from multiple sites (54 NICUs). However, these studies used a pooled, cross section of data from 1998-9, since which time the role of nursing in a neonatal unit is likely to have changed significantly; moreover, these studies were not able to examine within unit changes to nurse

to patient ratios as this chapter has done. All of the other relevant studies identified in the literature review (Callaghan et al., 2003; Cimiotti et al., 2006a; Filho et al., 2011; Grandi et al., 2010; Profit et al., 2010), also utilised a pooled, cross section of data, and often from a much smaller number of neonatal units.

The limitations of this study must also be acknowledged. This chapter has provided evidence that increases in the proportion of intensive care days that are one to one caredays leads to a reduction in the mortality rate. However, it is not possible to discern whether the one to one nurse to patient ratio is the optimal ratio for these infants. To identify the optimal ratio, we would require observations of a range of nurse to patient ratios and patient outcomes. Another question of interest is the effect of re-allocating nursing labour from other nursing tasks to one to one nursing (i.e. the magnitude of the parameter γ_0 in the conceptual model in Section 6.2). Without further information on the total amount of nursing labour available to each unit over time, it remains unknown how may additional nurse hours would be required to increase the provision of one to one nursing on neonatal units by a given amount. Recent time use data for nursing staff in the UK do not exist; the last such time use studies were conducted in the early 1990s and their results are unlikely to be generalisable to present day (Northern Neonatal Network, 1993; Williams et al., 1993). This therefore precludes cost-effectiveness analyses of any policy aimed at increasing one to one nursing provision. New time use studies are hence an important piece of future research.

As previously discussed, it is not possible to make inferences about the optimal nurse to patient ratio for infants receiving neonatal intensive care, nor is it possible to infer that nursing labour should be re-allocated to one to one care for neonatal intensive care from other nursing tasks. As the conceptual model in Section 6.2 showed, if nursing labour is reallocated to one to one nursing from other tasks, there is likely to be a negative impact on those patients who don't receive one to one nursing. Without further information on the total amount of nursing labour available to each unit over time, it remains unknown how may additional nurse hours would be required to increase the provision of one to one nursing on neonatal units by a given amount. Recent time use data for nursing staff in the UK do not exist; the last such time use studies were con-

ducted in the early 1990s and their results are unlikely to be generalisable to present day (Northern Neonatal Network, 1993; Williams et al., 1993). This therefore precludes cost-effectiveness analyses of any policy aimed at increasing one to one nursing provision. New time use studies are hence an important piece of future research.

This chapter does provide clear evidence to support the claim that an increase in one to one nursing provision on a neonatal unit reduces the risk of mortality. Further research is clearly warranted on the best way to achieve this. Furthermore, the benefits of increasing the nurse to patient ratio may be underestimated in this chapter since common neonatal morbidities are not considered, and previous studies have shown an increased nurse to patient ratio is associated with a reduction in the risk of diseases such as infection, BPD, and IVH (Cimiotti et al., 2006a; Grandi et al., 2010; Profit et al., 2010). This would be an important extension to these analyses in future.

In the previous chapter, it was shown that an increase in expenditure on neonatal healthcare led to a reduction in the risk of mortality for infants admitted to neonatal units at the place of birth. Moreover, it was shown that these increases were correlated with the nurse to patient ratio on neonatal units. We may tentatively argue that if funding was increased for neonatal units, at least part of the funding would be allocated to nursing labour, which should in turn increase nurse to patient ratios. Increased provision of one to one nursing may be one of the many mechanisms by which neonatal healthcare expenditure affects infant health. The results in this chapter, along with those of the previous chapter and the BAPM guidelines, provide evidence in support of increased nursing labour provision on neonatal units in England.

Chapter 7

Local Economic Conditions at Conception and Infant Health at Birth

The aim of this chapter is to investigate the link between local economic conditions at the time of an infant's conception and the health of that infant at birth. This topic is of interest in this thesis since it may be one of the mechanisms underlying the increase in admissions observed in Chapter 2. I examine infant health in England at a small area level in which changes to the local unemployment rate may be representative of broader local economic conditions and may therefore be correlated with changes to other factors, such as the local wage rate, which have a direct effect on fertility decisions, health behaviour, and subsequent infant health. Infant health at birth is measured in this chapter by the proportion of live births admitted to neonatal specialist care. While many studies have investigated infant health at birth previously, none have utilised an admissions rate to neonatal healthcare as a measure of infant health.

Both positive and negative relationships between local economic conditions and infant health have been documented in previous studies (for example, Dehejia and Lleras-Muney (2004); Lindo (2011)). These empirical studies have generally focussed on changes to birth weight as a measure of infant health at birth. However, birth weight is not a complete measure of infant health. Birth weight is correlated with health at birth but not perfectly. For example, an infant who is born at 28 weeks gestation and who weighs 1,500g (and so is classified as very low birth weight) is over four times

more likely to die while admitted to a neonatal unit than an otherwise identical infant born at the same weight but at 32 weeks gestation (Cole et al., 2010). Admissions to neonatal healthcare services, on the other hand, should be a direct function of overall infant health—admission decisions are made on the basis of the overall level of health, both observed and unobserved (to the analyst). The NNRD data utilised in this thesis reveal that only 43% of admissions to neonatal specialist healthcare services are of infants considered low birth weight (<2,500g). And, there is little evidence to suggest whether the changes to birth weight observed in studies of infant health at birth translate into a greater requirement for neonatal healthcare services which is an important concern for healthcare policy. This study examines the effect of local unemployment on the very low birth weight (<1,500g; VLBW) and low birth weight (<2,500g; LBW) birth rates alongside admission rates for comparison.

Economic theory predicts that, at a local area level, changes to economic conditions may affect aggregate infant health by altering household fertility decisions leading to a change in the overall composition of parents deciding to have children and affecting the health-promoting behaviours of those parents. Results from the empirical literature that motivate this study are inconclusive. Dehejia and Lleras-Muney (2004) found that, at the state level in the United States, the local unemployment rate and state level infant health, as measured by the proportion of VLBW and LBW births, were counter-cyclically related while the unemployment rate did not appear to have an effect on the overall birthrate. They found that the effect on the VLBW and LBW birth rate was driven by a combination of a change in the characteristics of mothers having children and a change in the health-related behaviours of those mothers. This may lead us to expect utilisation of neonatal specialist healthcare to decline during times of recession. However, utilising individual level data, Lindo (2011) found that husbands' job losses led to a reduction in infant birth weight which suggests infants born after paternal job losses are in worse health, although, paternal job losses may not lead to a change in paternal unemployment.

Use of the admissions rate as a measure of infant health at birth may further introduce an additional complicating factor for the analyses. Within the health economics

literature, it has been documented that an increase to healthcare capacity often causes an increase in demand for those healthcare services (Fuchs, 1978). This supply induced demand effect may similarly lead to increased utilisation in neonatal units, potentially counteracting the effect of changing infant health (Freedman, 2012; Fuchs, 1978). This possibility is considered in the analyses in this chapter.

The level of socio-economic deprivation in an area has been widely documented to be related with both the local economic conditions, through effects on labour market outcomes, as well as with the health of the individuals residing in that area (Cutler et al., 2011; Ioannides and Loury, 2004). These are often referred to as neighbourhood effects. Similarly, a number of previous studies have identified socio-economic gradients in healthcare utilisation often arising from disparities in health which are only partly explained by differences in health promoting behaviour (Morris et al., 2005; Richter et al., 2009; Stringhini et al., 2010). Decisions regarding fertility and health promoting behaviour may also differ depending on socio-economic status and local social norms, which have previously been shown to affect parental child rearing values, among other factors (Bowles, 1998; Bowles and Polanía-Reyes, 2012). This study therefore also explores how the effects of changes to local unemployment differ between areas of high and low socio-economic deprivation and within areas that differ in terms of workforce occupational composition.

This analysis may therefore play an important role in the understanding of socio-economic health disparities. There has been much evidence in support of the fetal origins hypothesis—that the nine months *in utero* are one of the most critical periods in shaping a person's future health trajectory (Almond and Currie, 2011). The measure of infant health used here reliably captures a broad range of neonatal conditions—the disaggregated analysis at the level of socio-economic status may provide supporting or contrary evidence to the fetal origins hypothesis.

The results in this chapter show that within area increases to the local unemployment rate at the time of conception lead to increases in the proportion of babies admitted to neonatal care. I argue that these are causal effects. Evidence is not found of an effect of local unemployment on VLBW or LBW birth rates, unlike in Dehejia and

Lleras-Muney (2004). The effect on admissions is different for areas depending on the level of socio-economic deprivation—there is only evidence of an effect of local economic conditions for areas in the bottom two quintiles by socio-economic deprivation. In addition, an effect is observed for 'working class' areas but not 'professional' areas (these terms are defined later). Only changes to the number of economically active individuals appears to affect the admissions rate and, among those, an effect is observed only for changes among males. These results are tested for robustness using a number of sensitivity checks.

These results are important for several reasons. This chapter shows that changes to labour market conditions at the time of conception affect infant health as measured by the subsequent admissions to neonatal health care. Recent studies, as well as the results presented in Chapter 5, have shown the beneficial effect of spending on neonatal health care on infant health and later life outcomes (Almond et al., 2010; Bharadwaj et al., 2013). Within the UK there has been strong pressure to reduce healthcare spending at the national level (Appleby, 2012). However, if infant health varies cyclically, and subsequent healthcare utilisation therefore increases, as this study suggests, then reductions to healthcare expenditure during times of poor economic conditions may exacerbate poor health among newborns. As Chapter 5 showed, reductions to neonatal healthcare expenditure, led to increases in the risk of mortality among infants admitted to neonatal units. There is also evidence of a strong socio-economic gradient in infant health and a differential response to within area changes to the unemployment rate. From a policy perspective, this indicates the importance of targeting interventions at areas of high socio-economic deprivation to improve infant health and reduce healthcare utilisation. This study may also contributes to the growing literature on the developmental origins of health.

The rest of the chapter is organised as follows. Section 7.1 provides a theoretical background and outline of neonatal healthcare services in England. Section 7.2 details the data sources, analysis sample, and variables. Section 7.3 explains the econometric specification and estimation. Section 7.4 provides the main results, Section 7.6 details extensions to the main analysis, and Section 7.7 concludes.

7.1 Theoretical Background

Economic theory predicts that local economic conditions at the time of conception may affect infant health at birth through two primary channels. Firstly, there may be changes to household fertility decisions such that those households that choose to have a child may differ from those that postpone their fertility decision; secondly, changes to household income may affect parental health behaviours which have a direct effect on infant health. Dehejia and Lleras-Muney (2004) found evidence of both effects in the US.

There is a large literature on the effects of labour market conditions on the fertility rate (see Hotz et al. (1997) for a review). Economic theory assumes fertility decisions are made at the household level and predicts that improvements in male labour market conditions should lead to an increase in the fertility rate. This is due to an increased demand for children resulting from increased household income, assuming children are normal goods and that females are primarily responsible for child rearing. Improvements in female labour market conditions are predicted to have opposing income and substitution effects. Time allocation decisions are often modelled as being a trade off between leisure, work in the market, and work at home (for example, Gronau (1977)). Raising children is a time intensive activity and is considered as work at home. Through the substitution effect, a wage decrease reduces the opportunity cost of time, increasing the amount of time allocated to work at home and therefore potentially increasing fertility. The income effect should have the opposite effect by reducing the demand for children. The empirical evidence is mixed and has found that transient decreases to the wage rate both increase (Heckman and Walker, 1990; Hotz et al., 1997; Jones et al., 2010; Jones and Tertilt, 2008) and decrease (Lindo, 2010; Lovenheim and Mumford, 2013) the fertility rate.

Health promoting behaviours are generally time intensive, thus a reduction in the opportunity cost of time should increase time allocated to health promoting behaviour. For example, increases to unemployment have been shown to lead to increases in prenatal care use (Menclova, 2012) while smoking and drinking among women have been

shown to be associated with economic downturns (Dehejia and Lleras-Muney, 2004). However, evidence also shows that husband's job losses have a deleterious effect on infant health (Lindo, 2011). This may be due to increases in stress or a deterioration in maternal nutrition which has been shown to be a highly important determinant of infant health (Almond and Mazumder, 2011). In addition, van den Berg et al. (2006) found evidence that individuals conceived in times of economic downturns had shorter life expectancies which may be the result of worse health at birth.

The previously discussed theoretical context provides some predictions about household level fertility and health behaviour decisions. It is important to consider that the effect of economic conditions on these decisions are unlikely to homogeneous between areas at a more aggregate level. Different socio-economic groups may, on average, make different decisions due to local social norms and social preferences which may differ broadly depending on social class or other socio-economic grouping (Bowles, 1998). The effect of changes to household income on health related good consumption, such as tobacco, may also change by socio-economic group—a number of studies have identified that lower socio-economic groups have higher price elasticity of demand for tobacco than other groups (for example, Thomas et al. (2008)). As a result, the effects of unemployment on infant health are also examined within areas by level of socio-economic deprivation and labour type.

The previous arguments also suggest that there may be a a potential difference between short-term, transient changes to earnings and long-term, permanent changes. In the former case, we may expect to see an increase in health promoting behaviour for the aforementioned reasons, without seeing a large decrease in the consumption of health related goods due to consumption smoothing (Gruber, 1997). This may be dependent on access to savings or credit. In contrast, long term unemployment is likely to lead to reductions in household consumption directly but also reduced future income due to deterioration of skills and reduced probability of employment. Indeed, babies born in areas of high socio-economic deprivation have been shown to have worse health at birth (Smith et al., 2007, 2009).

The theoretical background provided in this section could be used to identify testable

hypotheses. However, a number of mechanisms by which economic conditions may impact infant health at birth are identified, making even the expected sign of the effect ambiguous. Furthermore, results from previous studies cannot necessarily be generalised to the United Kingdom. The availability and levels of state financial support for the unemployed differ in the United Kingdom from the United States. The availability of financial support may impact household fertility decisions. In any case, this highlights the importance of this empirical research in understanding the wider effects of changes to local economic conditions on population health.

7.2 Data and Variables

The NNRD contains data at the individual baby level. However, an analysis of a cross-section at the baby level would be complicated by individual unobserved heterogeneity which would lead to inconsistent estimates of the effect of the local unemployment rate, and, the NNRD only contains observations of infants admitted to neonatal units which comprises only 8-9% of the population. As such, a panel of aggregated data is created for small areas in England.

The unit of observation is the Middle Super Output Area (MSOA). MSOAs, and their constituent lower super output areas (LSOAs), are described in detail in Section 2.3.1 in Chapter 2. The period for which the data set is constructed is 2007 to 2011. Data are available for 2006, however, only 96 of 170 neonatal units contributed their data at this time (Table 2.1 in Section 2). In addition, 2006 was the first year of the NNRD and as such there may be issues with data quality. As a result, 2007 to 2011 is chosen as the period of analysis—during this time the total number of units contributing data and providing permission for its use increased from 128 to 161, as shown in Table 2.1, Chapter 2. As a test of the robustness of the results, data from 2006 are included in the analysis sample.

7.2.1 Outcome variables

Admissions: The primary outcome under consideration in this study is the proportion of live births admitted to neonatal specialist care services. The number of admissions from mothers residing in each MSOA is recorded using the LSOAs in the database. Issues of selection and data completion owing to missing maternal postcode are discussed below.

Clinical outcomes: The internationally used ICD-10 medical classification codes for each infant are recorded within the NNRD.[1] Counts of the number of admissions with an ICD code in the following groups from each MSOA are recorded: congenital malformations of the circulatory system (ICD-10 codes Q20-Q28), haemorrhagic and haematological disorders of the foetus and newborn (P50-P60), respiratory and cardiovascular disorders specific to the neonatal period (P20-P29), transitory endocrine and metabolic disorders specific to the foetus and newborn (P70-P74), conditions involving the integument and temperature regulation of the foetus and newborn (P80-P83), and observation and examination (Z00-Z13).

Live births by birth weight: Data on the number of live births are obtained from the ONS for each MSOA. In addition, the number of live births by birth weight classification are obtained, the birth weight classifications are: less than 1,500g, greater than 4,500g, and 500g bins in between. The latter data are not publicly available.

Main explanatory variable

Unemployment rate: The unemployment rate used in this study is the claimant count rate which is obtained from the ONS. The claimant count rate is the proportion of working age individuals claiming Jobseeker's Allowance (JSA)—the out of work benefit. It is paid to individuals meeting the following criteria: aged over 18[2] but below the state pension age, not in full time education, able and available to work, and actively seeking work. The individual must work on average less than 16 hours per week and

[1] International Statistical Classification of Diseases and Related Health Problems, 10th revision. The full ICD-10 database is available at http://apps.who.int/classifications/icd10/browse/2010/en

[2] In England and Wales in 2012, 1.3% of all live births were to mothers aged under 18.

have less than £16,000 savings. The denominator population for each area is the number of individuals aged 15 to 64 since the population aged 18 to 64 was not available. An age specific unemployment rate was not utilised for two reasons, firstly, to aid comparability with the previous analysis of Dehejia and Lleras-Muney (2004), secondly, this analysis aims to explore the effect of local economic conditions on infant health at birth, and uses the unemployment rate as a proxy for this.

The JSA is a measure of the unemployment rate and considers only economically active individuals. Another out of work benefit called 'income support' is also available in England and has the same elegibility criteria as the JSA except that it is for economically inactive individuals such as those who cannot work due to disability, providing care, or being pregnant. In general, the distinction between the two benefits is determined by whether an individual is actively seeking work or not and so changes to local economic conditions may affect the number of JSA claimants as well as the number of income support claimants. The data also provide separate measures of male and female benefit claimants. However, each of these measures are highly collinear with the others (correlation coefficients > 0.9) meaning identification of separate effects by gender or economic status is generally not reliable. However, in extensions of the analysis unemployment by gender and economic activity is examined.

The JSA claimant count rate is used here in the absence of other measures of economic conditions. A potential weakness of this measure is that it is affected by policy changes that reclassify individuals as economically active or not without necessarily being related to changes in actual employment (Manning, 2009). Over the period of this study, no such policy changes were implemented.

7.3 Specification and Estimation

The aim of this chapter is to estimate the effect of changes to local economic conditions on infant health at birth. Local economic conditions are measured by the local claimant count rate and infant health is measured, in the main analysis, by the proportion of live births admitted to neonatal specialist healthcare services, however the

VLBW and LBW birth rates are also examined. A number of different issues need to be taken into account when considering the econometric specification for this analysis. As such, results from four alternative models will be presented, each model dealing with different issues that are identified here. Moreover, not all the issues discussed here are relevant to all the analyses, depending on the data source from which the relevant variables are derived.

The four models that are considered are:

Model 1: Pooled, cross section

Model 2: Panel, fixed-effect

Model 3: Panel, fixed-effect with correction for sample selection bias

Model 4: Panel, fixed-effect with corrections for sample selection bias and dependent variable measurement error.

The last two models are only relevant in the cases where the dependent variable is derived from the NNRD (namely admissions rates).

7.3.1 Data issues

The main outcome considered in this chapter is the proportion of live births admitted to neonatal healthcare. The source of this measure is the NNRD. There are two issues with these data that need to be taken into account when specifying the econometric model. In particular, not all neonatal units in England contribute their data to the NNRD which may lead to inconsistency in the estimators utilised in this chapter if those units are in areas that are systematically different from areas near contributing units. Given the relationship between organisational characteristics and unit quality, measured by infant mortality for example, it is possible that neonatal unit characteristics may be related to the decisions to adopt a centralised electronic reporting system and to permit analysis of the data. In addition, neonatal units may misreport a mother's residence so that the total count of admissions from a particular areas may be above or below the actual number of admissions from that area.

I consider the following reduced-form specification on which the models above are based:

$$y_{it}^{*} = \text{Unemp.}_{it} * \beta + x_{it}'\gamma + \alpha_i + \rho_t + u_{it} \tag{7.1}$$

where y_{it}^{*} is the infant health outcome for area i at time t, Unemp._{it} is the one year lagged local unemployment rate used to measure local unemployment at the time of conception,[3] x_{it} are exogenous area level variables including an intercept, α_i is area unobserved heterogeneity, ρ_t are year fixed effects and region specific time trends to capture compositional changes in population due to immigration and changes in labour supply at the regional level,[4] β is the parameter of interest to be estimated, and u_{it} is an i.i.d. random error. Standard errors are estimated by cluster bootstrap with 500 replications. Unless stated otherwise, "statistical significance" refers to a 5% significance level.

7.3.2 Model 1: Pooled Cross Section

Model 1 assumes that the area unobserved heterogeneity has zero variance such that there are no relevant unobserved differences between areas, so that $\alpha_i = \alpha$. In this model, x_{it} contains a number of variables derived from the 2011 census: the proportion of married households, the proportion of individuals reporting 'good health' or 'very good health', and the proportion of individuals with a university degree. These variables are only observed once during the panel and are therefore time invariant. This model is estimated by ordinary least squares (OLS), however, if the assumption of zero variance of unobserved heterogeneity is not correct, then the OLS estimator will be inconsistent.

[3] While the unemployment rate is measured at time $t-1$, the subscript at t is used for convenience.

[4] As the sample is large N (up to 6,781) and small T (T=5), area specific trends would result in an incidental parameters problem, region specific trends are included instead. England is divided into nine regions.

7.3.3 Model 2: Panel Fixed Effects

The OLS estimator of (7.1) is inconsistent if a model allowing unobserved between area differences is appropriate. It is assumed that α_i is an unobserved random variable that is potentially correlated with the included regressors. This model is often referred to as the fixed effects (FE) model. The within estimator is used to estimate this model.

7.3.4 Model 3: Selection Effects

Observation of the primary outcome variable for MSOA i at time t is determined by whether or not the neonatal unit to which the infants were admitted contributes its data to the NNRD at time t—the large majority of infants are admitted to their nearest neonatal unit at some point during their care.

Table 2.1 in Chapter 2 shows the characteristics of the neonatal units contributing data to the panel for each year. It is clear that the number of units increases substantially over the period 2007-11 from 128 to 165. Table 7.1 shows the differences between included/observed MSOAs and excluded/unobserved MSOAs. There is some evidence of a systematic difference between observed and unobserved areas in terms of the unemployment and birth rates. With the exception of 2010, unobserved areas have a higher unemployment rate than the observed areas and for all years unobserved areas have a lower birth rate. In the case where there is a systematic difference between areas, then the within area estimator used for model 2 will be inconsistent.

The presence of sample selection bias is tested for using two tests. Let s_{it} be an indicator equal to one if the MSOA i is observed at time t and zero otherwise. Nijman and Verbeek (1992) propose a simple variable addition test, namely adding $s_{i,t-1}$ as an explanatory variable to equation (7.1). Under the null hypothesis s_{it} is uncorrelated with u_{it} for all t, so $s_{i,t-1}$ should be non-significant. A robust t-test on this variable yields a p-value of <0.001. Wooldridge (1995) also suggests a test for selection effects in panel data: by estimating a selection model, determining the inverse Mill's ratio (IMR), including it in equation (7.1), and testing its significance (following Heckman (1979)). This method requires at least one instrumental variable for selection—here, I

Table 7.1 Summary statistics for included/observed and excluded/unobserved MSOAs.

Year	Variable	Observed	Unobserved	P-value
	n	4,777	2,004	
	Unemployment rate	2.2 (1.6)	2.5 (1.8)	< 0.000
2007	Birth rate	13.1 (4.6)	12.3 (4.2)	< 0.000
	VLBW birth rate	0.1 (0.2)	0.1 (0.2)	0.685
	Population (15-64)	4,999.3 (1,095.6)	4,886.8 (1,079.6)	< 0.000
	n	5,255	1,526	
	Unemployment rate	1.9 (1.5)	2.2 (1.7)	< 0.000
2008	Birth rate	13.2 (4.7)	12.7 (4.5)	< 0.000
	VLBW birth rate	0.2 (0.2)	0.2 (0.2)	0.339
	Population (15-64)	5,030.7 (1,131.8)	4,922.3 (1,099.6)	0.001
	n	5,568	1,213	
	Unemployment rate	2.1 (1.6)	2.3 (1.5)	< 0.000
2009	Birth rate	13.0 (4.7)	12.6 (4.5)	0.009
	VLBW birth rate	0.2 (0.2)	0.2 (0.2)	0.894
	Population (15-64)	5,048.5 (1,167.2)	4,976.0 (1,121.2)	0.043
	n	6,395	386	
	Unemployment rate	3.7 (2.1)	3.6 (2.0)	0.142
2010	Birth rate	13.2 (4.8)	12.3 (4.3)	< 0.000
	VLBW birth rate	0.2 (0.2)	0.2 (0.2)	0.351
	Population (15-64)	5,055.5 (1,200.1)	5,102.8 (1,197.2)	0.451
	n	6,332	449	
	Unemployment rate	3.3 (2.1)	3.1 (1.9)	0.015
2011	Birth rate	13.1 (4.7)	12.5 (4.5)	0.012
	VLBW birth rate	0.2 (0.2)	0.1 (0.2)	0.674
	Population (15-64)	5,084.5 (1,253.7)	5,081.4 (1,225.8)	0.958

[1] Values are mean(sd). P-values are from t-test of equality of means. Observed areas are those where the nearest neonatal unit contributes data to the National Neonatal Research Database in the given year.

[2] The unemployment rate is the JSA claimant count rate which is the percentage of the working age population claiming JSA.

[3] The birth rate is the number of live births per 1,000 population—this is the standard measure.

[4] VLBW = very low birth weight (<1,500g); The VLBW birth rate is the number of VLBW live births per 1,000 population.

use the proportion of units from the local neonatal network contributing in the previous year to the NNRD. This requires the assumption that neonatal units are more likely to contribute to the NNRD in this period if a greater proportion of units in their network were contributing in the last period. The variable is a statistically significant predictor of selection (p<0.001); it is estimated that a 10% increase in the proportion of units contributing in the local network last year led to a 6% increase in the probability of contributing data to the NNRD and hence being included in the sample. In addition, nearest neonatal unit designated care level is also included in the selection equation. For the purposes of testing, the selection equation is estimated as a pooled probit.[5] The p-value from a t-test of the coefficient on the included IMR is 0.079.

The preceding tests provide some evidence for the presence of sample selection bias. In equation (7.1) we must therefore condition on selection, the conditional expectation function (CEF) is therefore

$$E(y_{it}^{*}|\text{Unemp.}_{it}, x_{it}, \alpha_i, \rho_t, s_{it}=1) = \text{Unemp.}_{it}\beta + x_{it}'\gamma + \alpha_i + \rho_t + E(u_{it}|\text{Unemp.}_{it}, s_{it}=1) \tag{7.2}$$

A number of methods have been proposed in the literature to correct for selection in panel data models with unobserved heterogeneity. This analysis follows the method proposed by Wooldridge (1995). The selection equation is

$$s_{it} = 1(\delta_0 + z_{it}'\delta_1 + \delta_2 * \text{Unemp.}_{it} + x_{it}'\delta_3 + \xi_i + v_{it} > 0) \tag{7.3}$$

where z_{it} is the previously specified instrumental variable, v_{it} is an $N(0,1)$ random variable, and ξ_i is area unobserved heterogeneity. Estimation of model (7.2) then proceeds as follows. Letting $x_{it}^{+} = [z_{it}, \text{Unemp.}_{it}, x_{it}]'$:

1. Estimate equation (E.1) by standard probit separately for each time period, letting $\xi_i = \gamma_0 + \bar{x}_{it}^{+'}\gamma_1 + u_{3it}$, and obtain the inverse Mill's ratio, $\hat{\lambda}_{it}$. The inverse Mill's ratio is the ratio of the probability density function to the cumulative dis-

[5] This procedure is not a valid procedure for correcting for sample selection bias as it would be in a cross-section model (Wooldridge, 1995)

tribution function, $\hat{\lambda}_{it} = \frac{\phi(\hat{\delta}_0 + z'_{it}\hat{\delta}_1 + \hat{\delta}_2 * \text{Unemp.}_{it} + x'_{it}\hat{\delta}_3 + \xi_i)}{\Phi(\hat{\delta}_0 + z'_{it}\hat{\delta}_1 + \hat{\delta}_2 * \text{Unemp.}_{it} + x'_{it}\hat{\delta}_3 + \xi_i)}$

2. Model the unobserved heterogeneity in equation (7.1) as

$$\alpha_i = \theta_0 + \theta_1 \text{Unemp.} + x'_{it}\theta_2 + u_{\alpha,i} \tag{7.4}$$

and let $w_{it} = [x^+_{i1}, ..., x^+_{iT}, x_{it}, 0, ..., 0, \hat{\lambda}_{it}, 0, ..., 0]$ then estimate $y_{it} = w'_{it}\psi + \text{error}_{it}$ by pooled OLS.

7.3.5 Model 4: Measurement Error

The second NNRD data issue is considered here. For NNRD derived outcome variables, units may misreport or not report at all the location of the maternal residence. In the NNRD data, 6% of infants have a missing value for location of maternal residence (see Table 2.2 in Chapter 2). This may lead to inconsistency of the estimators of the previously described models if misreporting is correlated with the unemployment rate in a particular area. To see this, let e_{it} be the error in the NNRD dependent variables for area i at time t, and let y_{it} be the observed value of y^*_{it} such that $y_{it} = y^*_{it} + e_{it}$ (if there is no measurement error then $e_{it} = 0$).

As previously mentioned, measurement error is only an issue for the NNRD data, since the ONS data cover all MSOAs over the period of the panel. Moreover, it is assumed that the data obtained from the ONS are of a high standard with few or no errors. In the case where the same outcome is observed from both ONS and NNRD data, then it would be possible to obtain an estimate of e_{it}, as it is assumed that, for all ONS data $e_{it} = 0$. There is one variable that may satisfy this requirement—all VLBW live births should be admitted to neonatal care—the VLBW live birth rate should equal the VLBW admission rate.

Let $\hat{e}_{it}$ be the estimate of the measurement error, e_{it}, derived from comparisons of the VLBW birth rate in ONS and NNRD data (i.e. $\hat{e}_{it} = y^*_{it} - y_{it}$). It is important to recall that $\hat{e}_{it}$ is only observed for selected areas (i.e. $s_{it} = 1$), otherwise the CEF could contain e_{it} in equation (7.2). Replacing y^*_{it} with y_{it} in model 3 (equation (7.2)), I

specify the CEF as

$$E(y_{it}|\text{Unemp.}_{it}, x_{it}, \alpha_i, \rho_t, s_{it} = 1) = \text{Unemp.}_{it}\beta + x'_{it}\gamma + \alpha_i + \rho_t + E(u_{it}|\text{Unemp.}_{it}, x_{it}, \alpha_i, \rho_t, s_{it} = 1) - E(e_{it}|\text{Unemp.}_{it}, x_{it}, \alpha_i, \rho_t, s_{it} = 1) \quad (7.5)$$

If the error, e_{it}, were conditionally independent of the right hand side variables, so that $E(e_{it}|\text{Unemp.}_{it}, x_{it}, \alpha_i, \rho_t, s_{it} = 1) = 0$, then estimators of model 3 would be consistent even in the presence of this measurement error. Let the CEF for the measurement error be

$$E(e_{it}|\text{Unemp.}_{it}, x_{it}, \theta_i, \rho_t, s_{it} = 1) = \text{Unemp.}_{it}\pi_1 + x'_{it}\pi_2 + \theta_i + \rho_t + E(\varepsilon_{it}|\text{Unemp.}_{it}, x_{it}, \theta_i, \rho_t, s_{it} = 1) \quad (7.6)$$

where ε_{it} is an i.i.d. random error term. Then, substituting equation (7.6) into equation (7.5), gives the CEF for y_{it}

$$E(y_{it}|\text{Unemp.}_{it}, x_{it}, \theta_i, \rho_t, s_{it} = 1) = \text{Unemp.}_{it}(\beta - \pi_1) + x'_{it}(\gamma - \pi_2) + \theta_i + \rho_t + E(u_{it} + \varepsilon_{it}|\text{Unemp.}_{it}, \theta_i, \rho_t, s_{it} = 1) \quad (7.7)$$

and it is clear that the estimators of β will be biased by some amount π_1 equivalent to the (partial) correlation between the measurement error and the unemployment rate. This suggests a test for non-random measurement error in the dependent variable, H_0 : $\pi_1 = 0$; a robust t-test of this hypothesis gives a p-value of <0.001 thus we reject the null hypothesis of no measurement error.

Both equations (7.6) and (7.7) can be consistently estimated following the method in Wooldridge (1995) used to estimate model 3 and as described in Appendix E. As such estimates of β are obtained for model 4 by differencing the unemployment coefficients in equations (7.6) and (7.7), $\hat{\beta} = \hat{\tilde{\beta}} - \hat{\pi}_1$, where $\tilde{\beta} = \beta - pi_1$.

7.3.6 Other Issues

The unemployment rate is unlikely to be endogenous with respect to individual fertility decisions and health behaviour at the aggregate level once area fixed effects are included. And, since JSA is only available to those actively seeking work, it will not capture women who leave jobs in anticipation of pregnancy. However, the unemployment rate may capture the effects of coincident shocks to both economic conditions and infant health or be capturing the effect of an omitted variable. The model 3 is also estimated using the lagged local unemployment rate as an instrument for unemployment following the method of Semykina and Wooldridge (2010). These are not presented as the main results since the interpretation of the estimated effect would change, moreover, the consistency of an IV estimator in the face of the issues described below is difficult to establish.

The analyses are conducted in R 3.0.1 and Stata version 13.

7.4 Results

7.4.1 Descriptive statistics

Table 7.2 shows summary statistics of the population including numbers of benefit claimants, births, and admissions to neonatal care. Figure 7.1 displays the trend in benefit claimants graphically. Both Table 7.2 and Figure 7.1 show an increase in total unemployment over the period of the panel. This increase is primarily attributable to increases in the number of JSA claimants and hence those who are economically active. While the total number of male JSA claimants is approximately double the number of female JSA claimants, there are large increases for both genders. The total number of income support claimants remains fairly stable over the period (approximately 1.6 million) of which around 65% are female.

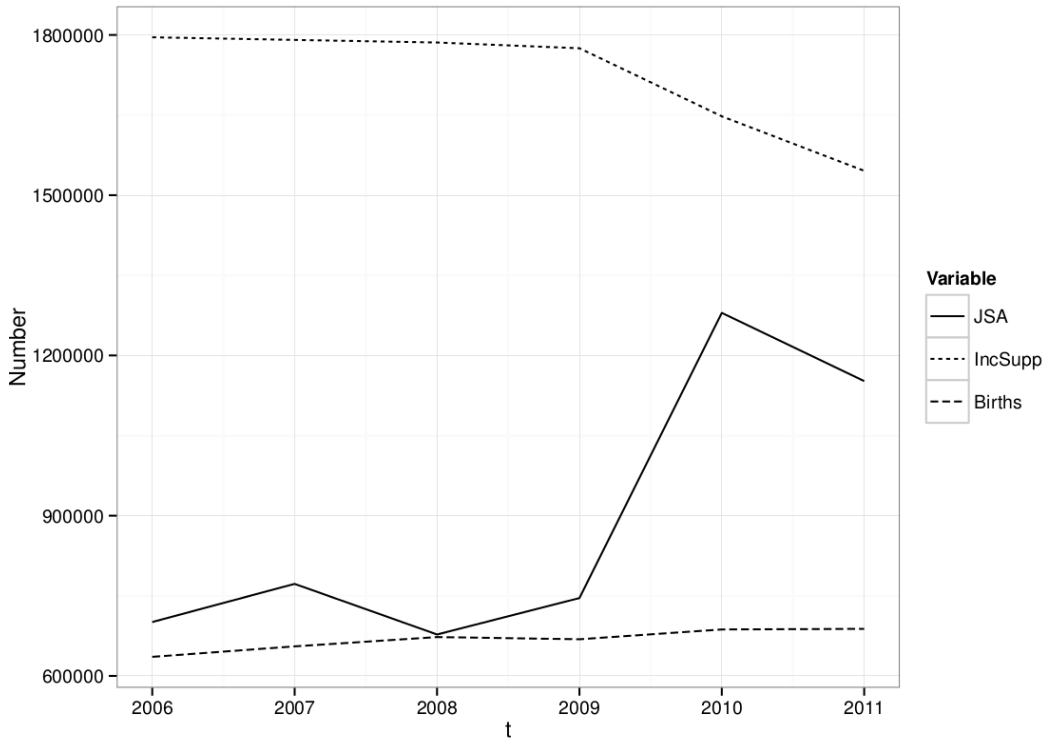
Table 7.2 shows that the number of JSA claimants increased from 772,315 in 2006 to 1,151,985 in 2011, an increase of 49.2%. The proportion of JSA claimants who were male was 72.3% in 2006 compared to 72.7% in 2011. The average observed

Table 7.2 Summary statistics for MSOAs by year

Variable	2007	2008	2009	2010	2011
Population[a]					
Population	50,763,893	51,106,181	51,464,646	51,809,741	52,234,045
Population 15-64	33,674,757 (66.3)	33,947,966 (66.4)	34,146,003 (66.3)	34,299,508 (66.2)	34,476,638 (66.0)
Population, Male	24,924,056 (49.1)	25,118,869 (49.2)	25,323,494 (49.2)	25,514,571 (49.2)	25,757,629 (49.3)
Population, Female	25,839,837 (50.9)	25,987,312 (50.8)	26,141,152 (50.8)	26,295,170 (50.8)	26,476,416 (50.7)
Lagged Unemployment[b]					
JSA, total	772,315 (1.5)	677,580 (1.3)	745,750 (1.4)	1,279,695 (2.5)	1,151,985 (2.2)
JSA, male	558,595 (2.2)	484,405 (1.9)	538,120 (2.1)	919,580 (3.6)	796,905 (3.1)
JSA, female	213,720 (0.8)	193,175 (0.7)	207,630 (0.8)	360,115 (1.4)	355,080 (1.3)
Inc Supp, Total	1,790,805 (3.5)	1,785,950 (3.5)	1,775,215 (3.4)	1,647,695 (3.2)	1,545,675 (3.0)
Inc Supp, Male	635,145 (2.5)	636,870 (2.5)	634,675 (2.5)	566,370 (2.2)	523,465 (2.0)
Inc Supp, Female	1,155,660 (4.5)	1,149,080 (4.4)	1,140,540 (4.4)	1,081,325 (4.1)	1,022,210 (3.9)
Births[c]					
Births	655,357	672,809	668,678	687,006	688,119
<1500g	7,482 (1.1)	8,044 (1.2)	7,936 (1.2)	8,171 (1.2)	7,982 (1.2)
1500-2000g	9,337 (1.4)	9,507 (1.4)	9,546 (1.4)	9,461 (1.4)	9,542 (1.4)
2000-2500g	29,954 (4.6)	30,282 (4.5)	30,294 (4.5)	30,067 (4.4)	30,816 (4.5)
>2500g	601,324 (91.8)	619,630 (92.1)	618,334 (92.5)	632,538 (92.1)	633,609 (92.1)
Admissions[d]					
Admissions	7.6	7.9	8.3	8.4	9.0
Mortality	1.9	2.0	1.7	1.7	1.6
IC	28.5	28.5	26.7	27.4	26.1
Congential malformations	6.4	7.2	6.8	7.0	6.7
Haematological	13.8	14.7	10.3	9.1	8.9
Respiratory and CV	39.6	43.2	42.6	43.1	42.2
Metabolic Conditions	19.7	20.5	21.1	21.7	22.3
Temperature Regulation	5.3	5.8	5.3	5.6	5.2
Observation	0.7	1.7	4.0	4.3	4.6
Birthweight	2623.1 (945.6)	2651.0 (953.4)	2701.2 (940.5)	2723.3 (933.1)	2750.3 (928.1)

[1] Values are n(%) unless otherwise stated.
[a] % as a proportion of the total population
[b] % as a proportion of total/male/female population aged 15-64.
[c] % as a proportion of total births.
[d] Mortality, IC, congenital malformations, haematological disorders, respiratory and CV are proportions of births and only include infants admitted to neonatal care.

Fig. 7.1 Total numbers of lagged JSA claimants, lagged income support claimants, and births between 2006-11 in England



change in unemployment for each MSOA was +14.5% or approximately one percentage point (pp). Table 7.2 also shows the proportion of admissions reported as experiencing certain conditions. The proportion of VLBW and LBW live births remained approximately constant at around 1.2% and 7.5% respectively over the course of the panel. The average birth weight of admitted live births increased by approximately 50g per annum. The NNRD data show that the proportion of admissions that were VLBW live births decreased from 14.0% in 2007 to 12.4% in 2008 and 10.0% in 2011.

The proportion of admissions is presented for each calendar year in the study period, 2007-11, which shows a large, 1.4 pp increase in admissions from 7.6% of infants in 2007 to 9.0% in 2011. These figures are calculated annually using the observed MSOAs in each year, but this increase may be attributable to not observing the admissions for every area. However, when considering only those areas observed in every time period, the proportion of live births admitted from these areas in 2011 is still 9.0%. Based on the total number of live births, this suggests that approximately 61,500 infants were admitted to neonatal care in 2011 up from approximately 49,800 in 2007,

Table 7.3 Effect of the local unemployment rate on admissions to neonatal care

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
Unemployment rate	0.168***	0.263***	0.412***	0.358***
	(0.017)	(0.0613)	(0.0671)	(0.0666)
Area fixed effect		X	X	X
Sample selection bias correction			X	X
Measurement error correction				X
N	28,542	28,542	28,542	28,542

[1] Cluster robust standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001

[2] The dependent variable in all regressions is the proportion of live births admitted to neonatal specialist care. The unemployment rate is the percentage of working age individuals claiming Job Seeker's Allowance (JSA). Model 1: pooled cross-section, model 2: panel fixed effects, model 3: panel fixed effects with corrections for sample selection bias, model 4: panel fixed effects with corrections for sample selection bias and measurement error in the dependent variable. The models are described in Section 7.3.

an increase of 23.5%.

As discussed earlier, the principal measure of infant health in this study is the proportion of live births admitted to neonatal units, this is compared to birth weight which is arguably the most widely used measure of infant health. The results examining the admissions rate are presented first, these are in Table 7.3, followed by birth weight results which are shown in Table 7.4. Following this, the effect of the local unemployment rate on various clinical outcomes of admitted infants are examined, and finally changes to the birth rate are examined.

7.4.2 Admissions

Table 7.3 examines the estimated effects of the unemployment rate on the percentage of live births admitted to neonatal specialist care. The estimated effect from all models is positive and statistically significant. The magnitude of the estimated effect is much larger in models 2-4. These results provide evidence of the presence of unobserved area effects which are not captured in model 1. Correction for sample selection bias in model 3 has the effect of increasing the magnitude of the coefficient when compared to model 2. However, there is little qualitative difference between the results from models

Table 7.4 Effect of unemployment on infant birth weight

Dep. var.	(1) % VLBW	(2) % VLBW	(3) % VLBW	(4) % LBW	(5) % LBW
Unemployment rate	0.075***	0.00318	0.0134	0.485***	-0.0590
	(0.0042)	(0.0172)	(0.0188)	(0.0116)	(0.0397)
Area fixed effects		X	X		X
Samp. sel. bias corr.			X		
Meas. err. corr.			X		
Data source	ONS	ONS	NNRD	ONS	ONS
Model	Model 1	Model 2	Model 4	Model 1	Model 2
N	33,905	33,905	28,542	33,905	33,905

[1] Cluster robust standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001

[2] The dependent variable is the proportion fo live births that are either very low birth weight (VLBW; <1,500g) or low birth weight (LBW; <2,500g) live births. The data source for the outcome variable is stated as either the Office of National Statistics (ONS) or the National Neonatal Research Database (NNRD), these are described in Section 7.2. The unemployment rate is the percentage of working age individuals claiming Job Seeker's Allowance (JSA). Model 1: pooled cross-section, model 2: panel fixed effects, model 4: panel fixed effects with corrections for sample selection bias and measurement error in the dependent variable. The models are described in Section 7.3.

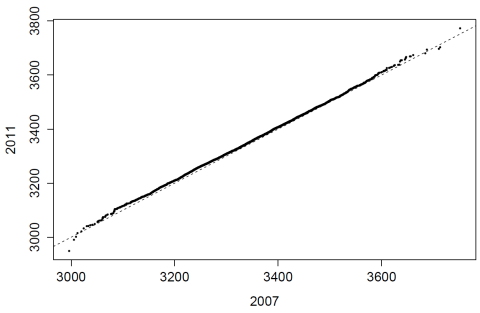
2 to 4. The main result in column (4), from model 4, suggests that a one percentage point increase in the unemployment rate would increase the admissions rate by 0.36 percentage points.

7.4.3 Birth-weight

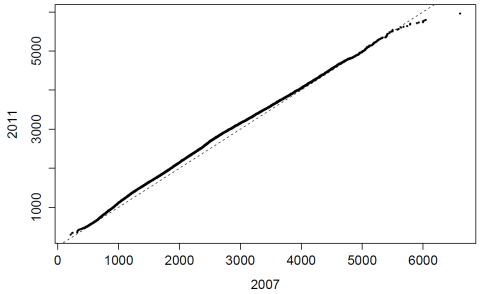
Table 7.4 shows the estimated effect of the unemployment rate on infant birth weight. The estimates of model 1 show a positive and statistically significant effect and imply that a one percentage point increase in the unemployment rate would result in a 0.04 and 0.38 percentage point increase in the VLBW and LBW birth rates respectively (columns (1) and (4)). However, the equivalent estimates of model 2 (taking into account unobserved heterogeneity between areas) are not statistically significant at the 5% level (columns (3) and (5)). Column (3) presents the results where the dependent variable is derived from the NNRD—this estimate is not statistically significant at the 5% level and is qualitatively similar to the result in column (2).

Birth weight is a common method of measuring baby health at birth. The dependent variables for the models examining birth weight are the proportion of infants out

Fig. 7.2 Birth weight quantile-quantile plots



(a) Quantile-quantile plot of average birth weights by MSOA for 2006 and 2011



(b) Quantile-quantile plot of birth weights of infants admitted to neonatal care in 2006 and 2011

of the total birth cohort whose birth weight falls below a certain threshold. Thus, considering the overall distribution of birth weights, these results suggest that either the unemployment rate did not affect the distribution of birth weights or shifted both the mean and variance of the birth weight distribution in the population so that the proportion of VLBW and LBW remained unchanged. It is not possible to definitively establish which of these cases is true, however, it is possible to examine the distribution of birth weights between two points in the panel.

Figure 7.2a shows a quantile-quantile plot comparing the distribution of average birth weight by MSOA for 2007 with the distribution in 2011,[6] this shows that the distributions were highly similar between the two years. Similarly, Figure 7.2b shows the distribution of birth weights for infants admitted to neonatal care in 2007 compared to 2011. This shows some increase in mass at 'normal' birth weights but only modestly.

7.4.4 Clinical outcomes

There are two potential explanations for the observed results. It is possible that aggregate infant health is deteriorating in response to increases in local unemployment but not through a mechanism that affects birth weight (hypothesis *a*). Or, alternatively the increase could be the result of supply side factors in some way related to the unemployment rate, such as increasing capacity or supply of physicians, unrelated to infant

[6]The average birth weight for each MSOA was calculated by multiplying the number of babies in each birth weight category by the mid-point of that category and dividing by the total number of births.

health (hypothesis *b*). To attempt to distinguish between these hypotheses, the clinical outcomes of those infants admitted to neonatal healthcare are examined next. In the case where the increase in admissions is unrelated to the health of the population of live births, hypothesis *b*, then the rate of adverse clinical outcomes observed on neonatal units among live births should not change in response to an increase in the local unemployment rate. However, if hypothesis *a* is true, then infant health at birth is deteriorating, and there should be a greater proportion of live births who experience adverse clinical outcomes on neonatal units.

Table 7.5 shows the effect of the local unemployment rate on the proportion of live births admitted with various conditions as well as the proportion of admissions dying while admitted to a neonatal unit and the proportion who receive any intensive care. Column (2) shows the estimated effect of the unemployment rate on the proportion of infants being admitted to neonatal care and dying while admitted to that neonatal unit, the coefficient is negative and statistically significant and suggests that a one percentage point increase in the unemployment rate is associated with a 0.04 percentage point reduction in the proportion of live births being admitted and dying. This is compared to a sample mean of 0.13% of all infants being admitted and dying. This result suggests that the risk of mortality among admitted infants reduces in response to an increase in the local unemployment rate which may be indicative of an increase in admissions of healthier infants (in support of hypothesis *b*). However, there are positive and statistically significant coefficients in columns (3) to (7), and (9). This would suggest that while these infants are at less risk of mortality, they are still unhealthy and require medical attention which provides evidence against hypothesis *b* and in support of hypothesis *a*. These results could plausibly be explained by an increase in average birth weight—low birth weight is a strong risk factor for mortality among babies (Medlock et al., 2011).

Table 7.5 Effect of local unemployment rate on the proportion of births admitted to neonatal healthcare with various conditions

	(1) Admissions	(2) Mortality	(3) IC	(4) Con. Mal.	(5) Haem. Dis.	(6) Resp. CV.	(7) Met. Con.	(8) Temp. Reg	(9) Obs. and Ex.
Unemployment rate	0.358***	-0.0418**	0.0871**	0.0342*	0.107***	0.187***	0.105***	-0.00743	0.0985**
	(0.0673)	(0.0144)	(0.0269)	(0.0172)	(0.0197)	(0.0311)	(0.0249)	(0.0172)	(0.0368)
Mean proportion (%)	6.97	0.13	1.88	0.51	0.80	2.99	1.46	0.37	2.73
Observations	28,327	28,327	28,327	28,327	28,327	28,327	28,327	28,327	28,327

[1] Cluster bootstrap standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

[2] These are results from model 4, described in Section 7.3.

[3] The numerator for the outcome variable in is the number of infants who were admitted to a neonatal unit and died ('mortality'), received at least one day of intensive care ('IC'; the most intensive level of care), or were recorded with an ICD-10 code in one of the following categories: Con. Mal. = congenital malformations, Haem. Dis. = haemorrhagic and haematological disorders, Resp. CV. = respiratory and cardiovascular disorders, Temp. Reg. = temperature regulation disorders, and Obs. and Ex. = observations and examination. The denominator is the total number of live births. The mean proportion is the arithmetic average of the dependent variable in the sample. The unemployment rate is the percentage of working age individuals claiming Job Seeker's Allowance (JSA). 7.3.

Table 7.6 Effect of unemployment on the birth rate

	(1)	(2)
	Model 1	Model 2
Unemployment rate	0.181***	0.207***
	(0.0209)	(0.0214)
Observations	33,905	33,905

[1] Cluster robust standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001

[2] The dependent variable is the number of live births per 1,000 individuals. The unemployment rate is the percentage of working age individuals claiming Job Seeker's Allowance (JSA). Model 1: Pooled cross section; model 2: panel fixed effects. The models are described in Section 7.3.

7.4.5 Birth rate

Finally, the effect of the local unemployment rate on the birth rate is examined (measured here by the number of live births per 1,000 population, which is the standard measure used in demographic studies), these results are shown in Table 7.6. While local unemployment may increase the proportion of live births admitted to neonatal care, the total number of live births may decline leading to an ambiguous change to the total number of admission to neonatal care. This is important for healthcare policy. Table 7.2 shows that the total number of live births in England increased by 33,000 between 2007 and 2011 and the proportion of admissions also increased. As Table 7.6 shows, the coefficients from both Model 1 and Model 2 are qualitatively similar and suggest that a one percentage point increase in the local unemployment rate would lead to an increase of 0.18 and 0.21 births per 1,000 individuals, respectively. Both results are statistically significant. This shows that local unemployment leads to an increase in the number of admissions both by increasing the size of the birth cohort and worsening infant health.

7.5 Robustness and Sensitivity

In this section, I provide results from a number of robustness and sensitivity tests.

Table 7.7 Effect of local unemployment on admissions to neonatal care, various methods for selecting observed MSOAs.

	(1) All areas	(2) No correct.	(3) Exact match	(4) All obs.	(5) Within 25%	(6) Incl. 2006
unemp	0.456***	0.718***	0.261*	0.263***	0.277**	0.412***
	(7.16)	(10.11)	(2.54)	(3.63)	(3.02)	(5.96)
Observations	33,905	28,542	18,131	24,025	19,016	28,542

[1] Cluster bootstrap standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001
[2] Results from Model 4, described in Section 7.3.
[3] All areas = all MSOAs included, No correct. = only selected areas but with no selection correction, Exact match.= only MSOAs where the VLBW count from ONS and NNRD match, All obs = only MSOAs that are observed in every time period, within 25% = MSOAs where the NNRD VLBW birth count is within 25% of the ONS VLBW birth count, and Incl 2006 = including 2006

7.5.1 Sample selection

The results are tested for robustness by altering the sample used in the estimation. These results are shown in Table 7.7. Column (1) shows the results using all MSOAs ignoring the sample selection problem (so that some areas will have a count zero admissions since the data are not in the database), column (2) displays results from the model with no selection correction, column (3) displays results using only MSOAs where the VLBW count matched from ONS and NNRD data sources. In column (4) estimates are presented using only areas that were observed in every time period, and column (5) displays results using a sample of MSOAs where the NNRD VLBW birth count was within 25% of the ONS VLBW birth count. Finally, data from 2006 were available to us, but were excluded since they differed significantly from the sample—only 43 units (of over 170) contributed data in this, the first year of the NNRD. The models are re-estimated, including 2006 and are shown in column (1). In all cases the estimated coefficient is both positive and statistically significant and similar in magnitude, or indeed slightly larger, than that estimated using our preferred sample. This provides evidence that the results are robust to method of sample selection.

Table 7.8 Effect of local unemployment on admissions to neonatal care with lagged local unemployment rate as instrument for local unemployment rate.

	(1) Pooled 2SLS Model 1	(2) FE-2SLS Model 2	(3) S&W Model 3
Unemployment rate	-0.0335	0.532	1.426***
	(0.0312)	(0.299)	(0.282)
Observations	23,260	23,737	23,737

[1] Cluster bootstrap standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001.
[2] The results are the intrumental variables counterparts to Model 1 (pooled cross section), Model 2 (panel fixed effects), and Model 3 (panel fixed effects with correction for sample selection bias) as described in Section 7.3. The estimators are, respectively, pooled two stage least squares (Pooled-2SLS), fixed effect two stage least squares (FE-2SLS), and the method proposed in Semykina and Wooldridge (2010) (S&W).
[3] The dependent variable is the percentage of live births admitted to neonatal care. The unemployment rate is the percentage of working age individuals claiming Job Seeker's Allowance (JSA).

7.5.2 Instrumental variables results

As a final robustness check, one year lagged local unemployment rate is used as an instrument for the local unemployment rate. These results are presented in Table 7.8. Instrumental variables estimates can be interpreted here as an average partial derivative of infant health with respect to the unemployment rate weighted in proportion to the instrument induced change in local unemployment rate. This is not necessarily the effect that this study intends to identify but any differences between this estimate and the estimates presented in preceding sections would provide information about the nature of the relationship between local unemployment rates. Furthermore, the instrumented results can be considered a causal effect (or weighted average of causal effects), the sign and statistical significance should provide support for the results in preceding sections.

The effect of instrumenting is to increase the magnitude of the estimated effect but the direction of the effect and statistical significance is unchanged. It may be inferred from these results that there are unobserved exogenous shocks increasing the local unemployment rate at the time of conception and decreasing the admissions rate or vice versa. Alternatively, those areas for which the effect of a change in the local

unemployment rate is greater in magnitude may be more highly weighted in the instrumental variable estimator. This may suggest non-linearities in the response function which suggests an avenue for future research. In any case, the conclusions of this study remain unchanged.

7.6 Extensions

7.6.1 Results by socio-economic status

This section examines MSOAs separately based on the level of socio-economic deprivation to explore whether there is a socio-economic gradient in within area effect. As discussed in Section 7.1, the relationship between local economic conditions and infant health at birth at an aggregate level is mediated through a number of different channels. The effect observed in this chapter is driven by a combination of various factors including fertility decisions, health behaviour decisions, and consumption decisions (Dehejia and Lleras-Muney, 2004). Even for just one of these factors it is highly unlikely that, conditional on employment status or income, households and areas would be homogeneous. The analysis is therefore disaggregated in two ways, by a measure of socio-economic deprivation and by a crude measure of class.

Index of Multiple Deprivation

Socio-economic deprivation is measured using the Index of Multiple Deprivation (IMD) 2010[7] which provides a relative measure of deprivation based upon seven socio-economic dimensions (income, employment, health, education, housing, crime, and environment) (Noble et al., 2007). Each LSOA, a smaller geographical unit than the MSOA, is assigned a rank according to the IMD. The IMD ranking for each MSOA is calculated by taking the harmonic mean of the ranks of the LSOAs which comprise each MSOA, weighted by the population in each LSOA. The harmonic mean is the reciprocal of the

[7]This index was created after the start of the panel but it derived from a number of variables in the 2001 Census. A prior Index was created in 2007 (Noble et al., 2007), in which the rankings of areas are very similar.

average of the reciprocals. The harmonic mean is used to aggregate rankings elsewhere (Chang and McAleer, 2013). The MSOAs are divided into quintiles based upon rank (i.e. assigned a quintile dummy equal to one if the MSOA appears in that quintile and zero otherwise), IMD quntiles are frequently used as measures of deprivation (Payne and Able, 2012).

Panel A in Table 7.9 shows the estimated coefficients from the model estimated including interactions of IMD quintile dummies with the unemployment rate. There is clear evidence of a socio-economic gradient in the effect of local unemployment shown in Table 7.9. There is evidence of an increased birth rate in response to an increase in the local unemployment rate in the top four quintiles of socioeconomic deprivation. There is no evidence of a change in the proportion of VLBW or LBW live births from any deprivation quintile. The pattern of results presented in the main results is repeated here for the most deprived areas but not for the least deprived quintiles. In particular, there is strong evidence of an increase in admissions for the top three quintiles (column (4)), the magnitude of the estimated coefficient is very similar to that estimated with the aggregated data in Table 7.4. We do not observe robust evidence of an increased proportion of babies admitted to neonatal care and receiving intensive care. However it is notable that column (6) shows a positive coefficient for the top three quintiles and a negative coefficient for the two least deprived quintiles. There is evidence of an increase proportion of live births being admitted and being diagnosed with haematological (column (8)) or respiratory and cardiovascular disorders (column (9)) but a reduced proportion of live births being admitted and dying (column (5)) for the most deprived areas.

Table 7.10 shows summary statistics of the top and bottom quintile areas by IMD in this study, supplemented using data from the 2011 census. The most deprived areas had a smaller proportion married households (average of 26.1% compared to 43.6% in the bottom quintile; $p<0.001$), a smaller proportion of individuals with a university degree or equivalent (9.7% vs. 19.4%; $p<0.001$), and a smaller proportion of individuals reporting 'good' or 'very good' health (77.5% vs. 86.1%, $p<0.001$). As the IMD reflects, socioeconomic deprivation is determined by a number of domains each of

which may contribute to the effects of changing unemployment on infant health.

Social class

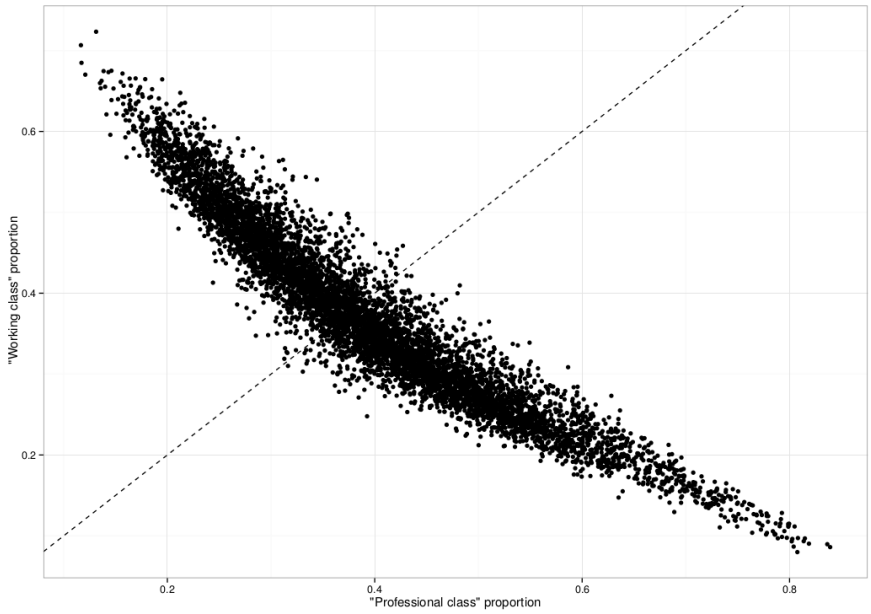
To examine the effects of local unemployment within social class, a crude categorisation of social class is constructed using the Standard Occupational Classification 2010 (SOC2010) (Office for National Statistics, 2010). Data are obtained from the ONS with the number of workers within various SOC2010 categories within each MSOA.

Various class schemas have been proposed in the literature; however, in general class schemas are used to classify individuals and households. Thus, an aggregation is required to classify areas. Within the UK, the National Statistics Socio-economic Comparison (NS-SEC) is used for official statistics. But, to classify each individual, information is required on employment status, occupation, and labour relations. Since these data are not available, only a crude class classification can be used here. The three class framework within the NS-SEC defines the top class as "Higher managerial, administrative and professional occupations" (hereafter referred to as 'professional' class) and the lowest class as "Routine and manual occupations" ('working' class). These classes are operationalised here by defining the 'professional' class as SOC2010 major categories one to three[8] and the 'working' class as SOC2010 major categories six to nine.[9] These categories classify 77.1% of workers in England. Figure 7.3 shows the correlation between the proportion of individuals classified as 'professional' class and the proportion of individuals classified as 'working' class for each MSOA - there is clearly a strong negative correlation which is expected given that these categorisations are intended to represent different ends of a class spectrum. A very simple schema is used in this study—areas are classified on the basis of whether there is a greater proportion of of 'professional' or 'working' class individuals. It is acknowledged that this method is a crude representation of class, however it is intended to capture class based on labour relations in way that perhaps the IMD cannot. In any case, it serves to

[8]These categories are: 1) Managers, directors and senior officials, 2) Professional occupations, and 3) Associate professional and technical occupations.

[9]These categories are: 6) Caring, leisure and other service occupations, 7) Sales and customer service occupations, 8) Process, plant and machine operatives, and 9) Elementary occupations

Fig. 7.3 Comparison of the proportion of individuals in 'professional' occupations and the proportion of individuals in 'working class' occupations for each MSOA



corroborate or contradict evidence seen using the IMD classification.

Class dummies are interacted with the local unemployment rate and results are shown in panel B of Table 7.9. In both 'professional' and 'working' class areas, an increase to local unemployment is associated with an increased birth rate (column (1)). But, only in 'working class' areas do we observe an increase in the proportion of live births admitted to neonatal care and the proportion admitted with haematological (column (8)) or respiratory and cardiovascular disorders (column (9)). There is no evidence of an effect on birth-weight in either area.

7.6.2 Economic status

The unemployment rate in this study was defined as the JSA claimant count rate. Only economically active individuals can claim JSA, however local economic conditions may equally affect the rate of economic inactivity, and the activities of economically inactive individuals may have an important effect on aggregate infant health. Table 7.11 shows the estimated coefficients from a regression of proportion of infants admitted onto unemployment rate measured by a variety of different claimant count rates. Inclusion of the income support claimants in the measure of the unemployment rate

Table 7.9 Effect of local unemployment on infant health disaggrated by quintile of the Index of Multiple Deprivation and social class

	(1) Births	(2) <1,500g	(3) <2,500g	(4) Admissions	(5) Mortality	(6) IC	(7) Con. Mal.	(8) Haem. Dis.	(9) Resp. CV.
				Panel A. Results by quintile of socio-economic deprivation.					
Least deprived[a]	-0.0587	0.0390	0.232	0.00896	-0.0865	-0.100	-0.0474	0.109	-0.129
	(-0.88)	(0.62)	(1.55)	(0.04)	(-1.85)	(-0.99)	(-0.83)	(1.54)	(-1.04)
	0.108*	0.0289	0.00292	0.231	-0.0404	-0.0313	-0.00509	0.154**	-0.0239
	(2.37)	(0.65)	(0.03)	(1.40)	(-1.14)	(-0.44)	(-0.12)	(2.83)	(-0.26)
	0.267***	0.0287	-0.0251	0.356**	-0.0391	0.0223	0.00335	0.108**	0.0651
	(6.90)	(0.89)	(-0.32)	(2.66)	(-1.36)	(0.45)	(0.10)	(2.67)	(1.05)
	0.273***	0.0334	0.0586	0.439***	-0.0456*	0.0537	0.0307	0.117***	0.111*
	(9.03)	(1.32)	(0.99)	(4.78)	(-2.16)	(1.38)	(1.21)	(3.94)	(2.27)
Most deprived	0.170***	-0.000946	-0.0698	0.286***	-0.0396*	0.0478	0.0264	0.108***	0.141***
	(7.57)	(-0.05)	(-1.70)	(4.01)	(-2.55)	(1.75)	(1.50)	(5.02)	(4.04)
Observations	33,905	33,905	33,905	28,542	28,542	28,542	28,542	28,542	28,542
				Panel B. Results by social class.					
'Professional' class[a]	0.169***	-0.0352	-0.123	0.144	-0.0421	-0.0349	0.00508	0.0759*	-0.00745
	(5.50)	(-1.31)	(-1.90)	(1.50)	(-1.81)	(-0.81)	(0.19)	(2.29)	(-0.14)
'Working' class	0.207***	0.00307	-0.0592	0.217***	-0.0200	0.0266	0.0304	0.0645**	0.0849*
	(9.73)	(0.18)	(-1.45)	(3.38)	(-1.29)	(0.98)	(1.73)	(3.01)	(2.40)
Observations	33,905	33905	33905	28327	28327	28327	28327	28327	28327

[a] Interactions of deprivation quintile or social class dummies with the unemployment rate variable.

[1] Cluster bootstrap standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

[2] Columns (1)-(3) are estimates of Model 2, columns (4)-(9) are estimates from Model 4. The models are described in Section 7.3.

[3] The outcome variable in column (1) is the birth rate, equal to the number of births per 1,000 population. In column (2) and (3) the outcome variables are the percentage of live births born at less than 1,500g and 2,500g respectively. The the outcome variable in columns (5)-(9) is the percentage of live births who were admitted to a neonatal unit and died ('mortality'), received at least one day of intensive care ('IC'; the most intensive level of care), or were recorded with an ICD-10 code in one of the following categories: Con. Mal. = congenital malformations, Haem. Dis. = haemorrhagic and haematological disorders, or Resp. CV. = respiratory and cardiovascular disorders. The unemployment rate is the percentage of working age individuals claiming Job Seeker's Allowance (JSA) and is interacted with dummies for socio-economic deprivation quintile in Panel A or class dummies in Panel B. Socio-economic deprivation is determined by the rank of the Index of Multiple Deprivation (IMD). Social class is determined according to the whether a greater proportion of individuals are classified in Standard Occupational Classification 2010 (SOC2010) major categories one to three ('professional' class) or SOC2010 major categories six to nine ('working' class).

Table 7.10 Summary statistics for areas with high and low average unemployment in 2006-11 using data from census 2011.

Variable	**Most deprived**	**Least deprived**	**P-value**
Short-term unemp.	3.1 (1.0)	3.6 (0.9)	<0.001
Married	26.1 (6.2)	43.6 (4.9)	<0.001
Education	9.7 (7.9)	19.4 (6.9)	<0.001
Health	77.5 (3.9)	86.1 (2.6)	<0.001
Unemployment rate	5.9 (2.0)	1.3 (0.4)	<0.001
Annual change	1.3 (2.5)	-0.8 (1.3)	<0.001

[1] Values are mean(sd).

[2] P-values are from t-test of equality of means.

[3] Short-term unemployment = proportion of total unemployed last employed in the previous 3 months, married = proportion of households married, education = proportion of individuals with university degree of equivalent, health = proportion of individuals reporting 'good' or 'very good' health, unemployment rate = claimant count rate, annual change = average annual change in the claimant count rate

results (column (1)) in a qualitatively very similar coefficient estimate to that shown in column (6) of Table 7.3. When the unemployment rate is measured solely by the income support claimants a positive and statistically significant coefficient.However, given the high degree of correlation between the income support claimant count rate and the JSA claimant count rate, the positive coefficient in column (2) may be the result of omitted variable bias. Column (3) presents estimates from a model with both claimant count rates. In this model only the JSA claimant count rate is statistically significant. This suggests that the observed effect in the preceding analyses is due to individuals seeking work although there does not appear to be much variation in income support over the period of the panel (see Figure 7.1).

As previously discussed in Section 7.1, since this analysis is conducted at an aggregate level, inferences cannot be made about individual level effects. Changing levels of JSA claimants may be reflective of shorter term unemployment changes when compared to changes the rate of income support claimants; alternatively, economically active individuals may have greater expectations of finding employment (hence their active status), thus they may have greater expectations of future earnings which would impact upon fertility decisions. Furthermore, the reasons for economic inactivity such as current pregnancy or disability, may also preclude or limit an individual's ability to have children.

Table 7.11 Effect of local unemployment on admissions to neonatal care, different measures of unemployment rate

	(1)	(2)	(3)	(4)	(5)	(6)
	Proportion of live births admitted to neonatal specialist care					
Unemployment rate	0.393*** (0.0578)	0.214*** (0.0645)	0.461*** (0.0707)	0.541*** (0.0832)	0.0304 (0.148)	0.536*** (0.0854)
Claimants	JSA and inc. supp.	Inc. supp.	JSA	Male JSA	Female JSA	Male JSA
Unemployment rate			-0.131 (0.0811)			-0.283 (0.155)
Claimants			Inc. supp.			Female JSA
Observations	28,542	28,542	28,542	28,542	28,542	28,542

[1] Cluster robust standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001
[2] Results are from Model 4. Models are described in Section 7.3.
[3] The dependent variable in all regressions is the proportion of live births admitted to neonatal specialist care. The unemployment rate is varies between columns, the denominator in all cases is the population aged 15-64, the numerator is a claimant count which is detailed below each coefficient. JSA = Jobseeker's Allowance, Inc. Supp.= Income Support.

7.6.3 Gender

The effect of changes to unemployment by gender are presented in columns (4), (5), and (6) of Table 7.11. In both cases the unemployment rate is measured as the gender-specific JSA claimant count rate where the denominator is the total population aged 15 to 64. The estimated coefficients for males is statistically significant, positive, and greater in magnitude than that presented in column (1) of Table 7.4; whereas the female coefficient in column (5) is not statistically significant (at the 5% level). In a model with male and female JSA claimant count rates entering separately we see (column (6)) a positive, statistically significant coefficient for males but a statistically insignificant coefficient for females. The results suggest that a one percentage point increase in the local male unemployment rate would lead to a 0.54 percentage point increase in the admissions rate whereas the female unemployment rate is not expected to impact on the admissions rate. This finding would be consistent with an assumption that males are the primary source of income in a household, these results would also appear to concur with those of Lindo (2011) who found, at the individual level, that paternal job losses during pregnancy led to a deterioration in infant health measured in that case by

birth weight. Furthermore, these results may also support the theory that females may substitute into health promoting behaviours which could be health protective for the baby.

7.7 Discussion and conclusions

This chapter has shown that between 2007-11 increases to the local unemployment rate at the time of conception led to an increase in the proportion of live births admitted to neonatal care (Table 7.3). There was no evidence of an effect on birth weight despite the strong relationship between the LBW birth rate and the rate of admissions (Table 7.4). In addition, it was found that these effects were only evident in deprived and/or 'working' class areas (Table 7.9) and that it was through effects on the economically active as opposed to the economically inactive and it acted through male JSA claimants as opposed to female (Table 7.11). These results were robust to a variety of checks including the addition of 2006 data and various specifications (tables 7.7 and 7.8). Furthermore, the correction for selection effects and measurement error do not change the general conclusions.

The mechanisms underlying the relationship between the economic conditions and infant are complex and not well understood. Previous research has shown there to be socio-economic gradients in health related behaviours such as smoking, drinking, and fruit and vegetable consumption. The link between smoking and birth weight has been well established (Hamilton, 2001; Juárez and Merlo, 2013), as has the link between maternal nutrition and birth weight Almond and Mazumder (2011). This study found evidence of increases in the proportion of live births admitted with various conditions of the newborn; however, the causes of these conditions are multifactorial and involve complex interactions of biological and environmental factors (March of Dimes, 2006). Despite this, some simple interventions, such as folic acid supplementation (De-Regil et al., 2010) have been shown to reduce the risk of newborn conditions such as those studied here. It has also been shown that elective caesarean section (c-section) is associated with increased risk of respiratory distress and neonatal care admission when

compared to vaginal birth and emergency c-section (Clark et al., 2010; Hansen et al., 2008)—the elective c-section rate has increased from 9.5% of all births in 2005-6 to 10.2% in 2011-12 (Information Centre, 2012) although it is unknown to what extent this may be related to local economic conditions. Further research is clearly required in order to link observed socio-economic phenomena with biological mechanisms.

The results presented here are important for a number of reasons. These results stand in contrast to the findings in Dehejia and Lleras-Muney (2004) who found that the local unemployment rate led to decreases in the low birth weight birth rate. The differences between this study and that of Dehejia and Lleras-Muney (2004) may be down to any of a number of differences: time period, country, level of aggregation, or measures of infant health. At the very least, these results suggest that birth weight may not a complete marker of infant health and may not be adequate on its own for evaluating aggregate infant health. As was discussed in the introduction, birth-weight may only be partially correlated with infant health. Alternative measures of infant health at birth may also be potentially useful for studies of this nature. The overall neonatal mortality rate along with the stillbirth rate would be notable candidates, however, these data were not available at the local area level. Nonetheless, the majority of neonatal deaths take place on neonatal units, the only exceptions are those deaths that take place prior to admission onto a neonatal unit. Developing aggregate measures of infant health at birth is an important topic for future research.

As argued in the introduction to this chapter, while birth weight is correlated with infant health at birth, it is not a perfect proxy. This chapter has shown discrepancies between the results where infant health at birth is measured by birth weight and those where it is measured by the admissions rate to neonatal healthcare. The admissions rate is arguably a function of infant health, capturing aspects of health at birth that are unobservable to the analyst. At the very least, this should be taken into account in future studies of this nature. Additionally, previous research has identified the effect of early life poor health on future outcomes (Almond and Currie, 2011; Currie et al., 2010), underscoring a mechanism for the intergenerational transmission of poor health, social and economic outcomes. Further research is required to elucidate the underlying

mechanisms. The results presented here may also aid planning and organising neonatal services at the national level

In conclusion, this study shows the wider implications of transient changes to labour market conditions, both in terms of increased healthcare utilisation but also that it may exacerbate socio-economic health inequalities. Effective policy needs to ensure that infants and mothers from more deprived areas receive adequate support to counteract the potentially deleterious effects of unemployment.

Chapter 8

Normative Issues Relating to the Use of Empirical Research in Neonatal Healthcare Policy

The evidence presented in these chapters may be used to inform policy debates relating to newborn health. It was shown that infants born in hospitals with higher volume neonatal units were at less risk of mortality; higher expenditure on neonatal healthcare and increased provision of one to one nursing reduced the risk of mortality for newborns; and increases to the unemployment rate at the time of conception led to worse infant health at birth. This may provide evidence in support of certain policies aimed at improving neonatal clinical outcomes. However, there are a number of normative issues that must be considered when assimilating the empirical evidence presented here into a coherent policy framework; often these normative issues are amplified within a neonatal context where small changes to healthcare can have large consequences over the course of an entire life.

It is important to frame the relevant policy decisions in the appropriate institutional setting. In the United Kingdom, economic evaluations of new health technologies and, to a lesser extent, healthcare policies are used to inform the social choice of technology adoption or policy implementation. The role of economic evaluation in social choice can be construed in two ways (Claxton et al., 2011). Either economic evaluation in-

forms "those responsible for maximising the present value of the time stream of health subject to exogenous budget constraints in each period" or, alternatively, economic evaluation has the wider objective of maximising social welfare (Claxton et al., 2011). The former role is the one that economic evaluation has generally taken in the UK. In fulfilling this role, the authority responsible for health economic evaluations in England, the National Institute of Health and Care Excellence (NICE), aims to maximise the health returns to an exogenously given set of resources. The healthcare technology assessment agency, in this role, does not express a social welfare function, rather the health maximisation objectives set by a principle can be seen as a "partial social expression of some unknown underlying latent welfare function" (Claxton et al., 2011). Where the objective of economic evaluation is to maximise social welfare, the authority would express an explicit social welfare function. The following discussion applies to both situations.

The purpose of this chapter is to discuss the key issues that relate to the use of the evidence presented in this thesis for decision makers. Evaluation of a neonatal healthcare policy, such as the centralisation of care, requires us to account for the benefits and the costs related to the implementation of such a policy. Then, this can be compared to other policies available or to a predefined cost-effectiveness threshold to healthcare decision makers to decide whether or not to implement such a policy. However, in doing so, we must answer a number of key questions. With regards to the benefits, should we discount gains in the future? Discounting over the length of a life leads to large differences in the cost per life year gained compared to a case of no discounting. At a 3.5% discount rate, a year gained in 80 years is only worth 23 days for an individual today—say at the end of a individual's life who is born today such as the recipients of neonatal healthcare. This leads to the large differences in the estimates of the cost per life year gained in Chapter 5 from £9,700 with 3.5% discounting to £3,210 with no discounting, a difference of around 300%.

How should we deal with the lack of data regarding quality and length of life among neonates? This issue makes the benefits accruing to neonates potentially incommensurable with those accruing to adults. Indeed, even within the newborn patient popu-

lation, how do we compare the benefits accruing to one group of patients with those accruing to, or even with the burdens (which may include foregone benefits) imposed on, another? Finally, at the decision maker level, when it comes to comparing policies, should we favour neonates over other patient groups? The rest of this chapter aims to address these questions and considers the policy of neonatal healthcare centralisation as an example.

8.1 The Social Discount Rate

There is an ongoing methodological debate about how to account for future health effects in health economic evaluations. This debate is especially pertinent when considering interventions that affect newborns where health effects may occur many decades in the future. There are a number of different reasons provided for discounting and thus different rates are recommended. These rates will vary depending on whether we are conducting cost-benefit analysis (CBA) or cost-effectiveness analysis (CEA) (Gravelle and Smith, 2001). In CBA, health effects and costs are converted into the same metric which is generally in terms of present day monetary value. CEA, on the other hand, aims to determine an incremental cost-effectiveness ratio (ICER) where consequences are expressed in terms of natural or physical units and preference-based outcomes, e.g. QALYs, for cost-utility analysis. In both cases it must be decided how to transform future benefits into present values or quantities. NICE recommend a discount rate of 3.5% for both costs and benefits (National Institute for Health and Care Excellence, 2013). [1]

Health effects can be expressed as the consumption value of health—the amount of consumption at time t equivalent to one unit of health. In this case, the social discount rate can be decomposed into two components: a) the time preference rate as applied to cardinal utility, and b) differences in the marginal utility of consumption over time.[2] (Claxton et al., 2011; Cowen, 2001; Cowen and Parfit, 1992; Gravelle

[1]There is a further debate about whether benefits and costs should be discounted at the same rate. This is considered by Claxton et al. (2011) but is beyond the scope of this discussion.

[2]This discussion revolves only around the discounting of health benefits rather than costs. In the

and Smith, 2001). Importantly, in neither of these cases is a positive discount rate necessarily implied. When comparisons are made, not only intragenerationally, but also intergenerationally then a positive discount rate may indeed be incorrect.

8.1.1 Time Preference

The extent to which an individual prefers present benefits to those in the future is known as an individual's time preference. While an *individual* may have a positive time preference, whether rational or or not, this does not necessarily imply a positive *social* discount rate. This is most clearly evident when we make intergenerational comparisons, such as the case when considering policies that affect the neonatal population, which are policies that may affect as yet unborn people. Often a positive social time preference for health is taken as given (such as in Claxton et al. (2011); Gravelle and Smith (2001)). Yet, while I may prefer to be in better health now rather than later, infants who are as yet unborn have no such preference. They are not 'waiting' for these benefits to accrue. This is known as the No Waiting Argument, the stronger form of which states that we have zero time preference before birth. It makes little sense to argue that we should discount future benefits for as yet unborn infants because they would prefer to receive those benefits now. At the very least, the No Waiting Argument shows that a positive discount rate within a life does not imply a positive discount rate across lives and hence a positive social discount rate, this is the weaker form of the No Waiting Argument.

Cowen (2001) argues that the No Waiting Argument has further, stronger implications in that unless the inter- and intragenerational discount rates are equal, then we end up with erroneous conclusions, particularly if we wish to be consistent with the policies we adopt over time (let this be the time consistency requirement). To adapt Cowen's argument, consider two alternative healthcare policies, one which provides a benefit of an extra 10 years of life to someone who would otherwise die in 2035, but who will not be born until 2025, or a policy which provides an extra 11 years of life

case of costs, we could arguably add a third reason—that the budget constraint increases over time so that future costs are less important since they lead to less health foregone.

to someone who would otherwise die in 2035 and who is alive today. If we treat intragenerational discount rates as exceeding intergenerational discount rates, then we may end up choosing the former policy which is clearly the inferior of the two. Furthermore, this choice of policy would fail a time consistency test, since if we re-evaluate our choice in 2034 we will change our policy selection (Cowen, 1990). Despite this, allowing different inter- and intragenerational discount rates is advocated by some authors as a solution to this problem (e.g. Gravelle and Smith (2001); Lipscomb (1989)).

If the No Waiting Argument is accepted and time consistency is required, then we are forced to accept that the intergenerational and intragenerational discount rates must be equal (and zero). This may sound counter-intuitive when considered in the context of an individual life, but we are talking about time preference for cardinal utility rather than for goods and services. It may be argued that a positive rate of time preference arises simply because we prefer present goods to future goods due to their higher rate of marginal utility and that pure time preference across cardinal utility does not provide an independent reason to discount. Thus, even if we abandon a positive discount rate on the basis of time preference for cardinal utility, there may still be compelling reasons to accept a positive discount rate in terms of the marginal utility of wealth. However, as is argued below, even if we accept this argument for wealth, it is likely that it does not apply to the case of health. Indeed, health may be increasing in value over time.

8.1.2 Marginal Utility of Consumption

A positive social discount rate for health effects is implied if the consumption value of health is decreasing. In the case of a decreasing consumption value of health then one unit of health is worth less in the future than it is today which means that present health benefits should be preferred to future health benefits. However, as Gravelle and Smith (2001) and Claxton et al. (2011) argue, the consumption value of health is likely to be *increasing*. One reason for this, they argue, is that the marginal utility of consumption is likely to diminish faster than the marginal utility of health. Individuals will generally be wealthier in the future, not least due to the increasing marginal productivity of

technology, thus future consumption goods will be valued less at the margin. The welfare gain from better health is unlikely diminish in the same way, both because population health is expected to grow more slowly than population consumption and because, at the individual level, health will deteriorate with age, on average and, at the population level, there will be a greater number of older individuals. Thus, the value of health will increase in terms of consumption. Furthermore, many studies suggest that the value of a statistical life increases with income and will therefore increase over time (Costa and Kahn, 2004; Hall and Jones, 2007).

An increasing consumption value of health would imply a negative social discount rate (if the pure time preference for cardinal utility were zero). However, the previous discussion is of the consumption value of health *at the margin*. Market prices, and other sources of evidence such as the interest rate, are used to estimate the rate at which individuals are willing to trade off marginal pounds (sterling) over time, but many of the health effects considered in these analyses are large, such as mortality, and concern willingness to trade off infra-marginal pounds (i.e. pounds below the margin) for which data generally do not exist.[3] The effects of a mortality affecting policy are large for each individual involved since they concern the entire stock of health rather than one unit of health; economic evaluations such as CBA concern themselves with individual valuations and, unless the changes are small for each individual, are not strictly correct.

Studies of the value of a statistical life may arguably be used to convert loss of life into small marginal changes, for example, where individuals have the choice of a 'risk to life' (Schelling, 1987). However, these studies cannot be used to make inferences about the value of a statistical life for future generations for whom no such choice is available. Moreover, when we consider a public health system, such as that in the UK, the public health system has access to health producing technologies, such as high-volume neonatal units, that private individuals do not. Market prices may not reflect the social marginal rate of substitution for health producing goods and services—

[3]Becker et al. (2005) do provide estimates of the value of infra-marginal changes to longevity across countries between 1960 and 2000 by estimating willingness to pay for increases in survival below the margin.

wealthier future individuals may wish to purchase more healthcare technologies that private technologies will allow at any price. The government may therefore redistribute healthcare production to the future where it will be more valuable, but this cannot be known without comparisons of infra-marginal units of health.

As a final point, as was demonstrated throughout this thesis, infants born to individuals from more socio-economically deprived areas are likely to be in worse health that their counterparts born in less deprived areas (see Chapter 7). While future generations may be, on average, wealthier, not all individuals will necessarily be wealther. Indeed, those that are least likely to be wealthier are those that are most likely to be in poor health (whatever the direction of causality). Thus, arguments about the social discount rate for health that are based on increasing wealth or incomes, such as the diminishing marginal utility of consumption, may not be valid for those very individuals we are valuing health benefits for.

Taken together, these arguments suggest that it may not be correct to assume a positive social discount rate for health benefits. It is even possible, if the pure time preference is zero, for the discount rate to be negative. Examples are provided in many studies considering the discount rate where even small changes to the discount rate can make a large difference to the policies and technologies selected (see, for example, Claxton et al. (2011)). This is also evident from the analyses presented in this thesis.

8.2 Measuring benefits

The benefits accruing to various policies or technologies in different fields of healthcare must be commensurable if a decision maker is to be able to reasonably select between those difference policies or technologies. The standard measure used by NICE, and indeed other technology assessment agencies around the world, is the Quality Adjusted Life Year (QALY) (Culyer, 2010). Incremental benefits are expressed in the length of time gained weighted by the quality of that time. Typically, quality refers to a health related quality of life (HRQoL) rather than a more generic quality of life (QoL) or well-being measure. Large numbers of consistent and reliable measures of

HRQoL have been developed for adults, with a growing number for children. In a systematic review, Solans et al. (2008) identified over 100 measures of HRQoL for pediatric populations; however, fewer than 30 of these were designed for infants of less than 5 years, and, importantly, none of these were preference based. The first measure designed specifically for infants under one year is the Peds QL Infant Scale, published in 2010 which is designed for children aged 1 to 24 months (Varni et al., 2011). Nonetheless, there does not exist any comparable measure that can be reliably applied to infants receiving care on the neonatal unit (Boss et al., 2012).

In general, economic evaluations within infant or neonatal populations, estimate the projected stream of health benefits of a particular treatment. However, this may not be a fair comparison with other areas of healthcare where *current* benefits are estimated. The issue lies with the difficulty of developing a measure of HRQoL for neonates; infants without linguistic and cognitive skills are unable to self-report quality of life, and, given the diverse developmental and emotional stages that underlie various disease states in children, adult HRQoL measures cannot simply be translated for an infant population. Boss et al. (2012) argue that a HRQoL measure can be developed for infants receiving care on a neonatal unit, taking into account many different dimensions of health already routinely measured as part of neonatal care such as pain metrics, neuro-behavioural metrics, and physical symptoms. Nonetheless, such a measure does not yet exist, and we rely simply on determining benefits in terms of changes to, for example, the mortality rate. This may mean we underestimate the benefits of various policies where there are improvements to the HRQoL of an infant on a neonatal unit but where the risk of mortality is otherwise unchanged, particularly among those admissions with a low risk of mortality. This may have a profound effect on the evaluation of non-Pareto efficient policies (those which benefit some at a cost to others) which have differential effects within the patient population. A good example of such a policy is centralisation where the evidence (see Chapter 4) suggests that only a small part of the patient population will benefit. Section 8.4 considers such a policy.

8.3 Distributional Justice

Finally, it is worth introducing a brief discussion of issues of distributional justice, and how alternative theories of how benefits ought to be distributed in the population may affect how interventions within the neonatal population are evaluated. While the role of economic evaluation may in practice be to inform healthcare decision makers rather than to explicitly maximise social welfare, it is important to consider whether the value of a QALY (or any other unit of health) is the same regardless of who receives it. The practice in the United Kingdom is generally to value all incremental health benefits the same, irrespective of the identity of the recipient or their current status, although end of life treatments are often given alternative weights.

The policy of treating all health benefits equally may be construed as maximising a utilitarian social welfare function under a budget constraint where the total benefit of a policy is simply the sum of all individual benefits. However, a number of authors have questioned, from a wide variety of different perspectives, whether 'a QALY is a QALY is a QALY' (Dolan and Tsuchiya, 2006).

The arguments against treating all QALYs equally can be categorised in two ways: i) utilitarian or libertarian, or ii) distributional.[4]

Valuing all QALYs equally is in contrast to the utilitarian approach behind CBA, where the value of a unit of health is derived from the willingness to pay (WTP) for that unit of health. From a libertarian point of view, if the distribution of income is fair, then incomes reflect contribution to society in which case QALYs should be weighted in favour of the WTP for each QALY. In this case the wealthy would be weighted over the poor. However, when considering the case of infants, who are completely dependent on others, the children of the wealthy would be weighted more highly that the children of the poor which very few people would accept—particularly if society places any value on equality of opportunity or democratic equality.

In the case where the distributions of income are not fair then another metric may

[4]Dolan and Tsuchiya (2006) classify the arguments against equal QALY weights as either: i) efficiency, ii) vertical equity, and iii) horizontal equity. For the purposes of this discussion, an alternative categorisation is more useful.

be used to differentiate individuals on the basis of their value to society, such as age. Age arguably reflects an individual's net contribution to society—infants are fully dependent on others but become more productive as they age until old age when they once again become dependent. The World Bank has produced estimates of age based efficiency weights (Murray, 1994). It is unclear whether these weights would affect assessments of neonatal technologies—if a policy allows an infant who would have otherwise died to live an otherwise normal life then the total value of the benefits remains unchanged as the total sum of the age based weights is the same as an equal QALY weighting, whereas if a policy only provides a small number of years of life into childhood then we may end up preferring policies that benefit adolescents or young adults. It is clear that data regarding lifetime outcomes are required.

The preceding discussions of utilitarian arguments against equal QALY weighting assume that societal value is measurable as a net fiscal contribution—an individual's net fiscal contribution has constitutive or intrinsic value. If markets are perfectly competitive then this may be a defensible claim. However, in the more likely case where the distribution of income is not fair, then net fiscal contribution may only be of instrumental value. There are many other ends that have been argued are the goal of policy, such as well-being or opportunity, that could be argued to promoted by maximising government revenues or economic growth. In this way, maximising efficiency is better for all since it generates revenues that can be used to improve healthcare for all. Efficiency based arguments may, however, lead to conclusions that many would reject. Those with higher incomes have better access to goods complementary to healthcare, such as healthy diets and exercise, that may make their use of public healthcare more efficient. Similarly, it may be more cost-effective to treat the healthy over the sick. This may then lead to an exacerbation of income or health inequalities to which there is an explicit societal aversion (Bobinac et al., 2012; Nord et al., 2010; Shah, 2009). In general, efficiency based arguments alone do not support weighting in favour of neonates, and may even lead to a weighting against neonates, at least for short-term interventions. However, if a concern exists for issues of equity then these conclusions may not hold.

Equal QALY weighting may be objected to on the grounds of resulting distribution of health or income in the population. Studies have revealed public preferences for more equal distributions of health in the population (Bobinac et al., 2012; Nord et al., 2010; Shah, 2009), and, as the preceding paragraphs discuss equal QALY weighting or efficiency based weighting may not achieve this. It is possible to identify various distribution-concerning normative theories that would support a non-equal QALY weighting; these include, but are not limited to: egalitarianism, prioritarianism, the maximin principle, luck egalitarianism, and democratic equality. Here, I provide a brief summary of these theories, and how they may favour or oppose benefits accruing to neonates.

Egalitarianism is the position that it is in itself bad if some people are worse off than others.[5] An egalitarian QALY weighting would therefore aim to promote an equal level of health in the population (whatever that level were). However, this view is subject to what is known as the Levelling Down Objection—we could, in effect, make the healthy part of the population sicker until everyone was equally as unhealthy. The Levelling Down Objection leads some authors to instead prefer Prioritarianism (Parfit, 1997).

Prioritarianism (also known as the Priority View) is not an egalitarian philosophy, rather it is an aggregative utilitarian philosophy, that posits that the total goodness of an outcome is the sum of all individual benefits with extra weight given to worse-off individuals. This view would promote an egalitarian outcome but does not hold that it is the resulting distribution of outcomes that determines the goodness of a policy. As Parfit (1997) writes "[O]n the priority view, we do not believe in inequality. [...] We do of course think it is bad that some people are worse off. But what is bad not that these people are worse off than others. It is rather that they are worse off than they might have been." One difficulty with this view, particular when it comes to operationalising it for use in QALY weighting, is determining the weights to be used. For example, in the most extreme case where all the weight is on the worse off with no weight given to

[5]There are different forms of egalitarianism, for example, Parfit (1997) identifies telic and deontic egalitarianism. The egalitarianism I focus on here is classified as telic egalitarianism under this distinction.

anyone else, we end up with the Rawlsian maximin principle (Parfit, 1997).

The quote in the preceding paragraph serves to emphasize that, under the Priority View, the concern is with absolute welfare rather than relative welfare. An infant who is in poor health and at risk of dying is in poor health even if there were no other infants that require medical assistance. It does not matter whether this infant is worse off than other infants. It is important to clarify what is meant specifically by 'badly off' in order to evaluate potential weighting schemes. The crucial point—particularly for interventions at the beginning of life—is that it is welfare over the course of a life that matters not just up until discharge from hospital or in the following few weeks. Consider that there exist policies that may reduce the mortality rate on the neonatal unit. Since the opportunity exists to reduce the mortality rate, such as by centralising neonatal healthcare, by not enacting such a policy we are allowing certain infants to die. The question is then whether it is wrong to allow these infants to die. It depends on which infants it is that we are considering; for example, many may agree that allowing an infant to die who would otherwise go on to live a small number of years into their childhood with a debilitating disability before dying would not necessarily be wrong. This suggests that reducing the mortality rate or keeping an infant alive is not in itself intrinsically good. However, we may argue that we do have a moral obligation in certain circumstances not to allow an infant to die—we ought to not allow an infant to die who would go on and live a full and healthy life. This would suggest that the value in reducing the mortality rate is instrumental in what it increases, whether that be welfare, quality of life, or capability to participate in democratic life. It is therefore clear that, here, whole life welfare should be the relevant basis for evaluation.

This discussion leads back to the question of the appropriate social discount rate if whole life welfare is to be taken into account. Different theories and discount rates would lead to different policies being accepted. As a simple example, consider the following scenario where we can enact either policy *A* or *B* where we can either provide a small benefit to a large number of people or a large benefit to one person::

Both policies A and B are arguably better than doing nothing at all.[6] In the case of

[6]'Strong egalitarians' may argue that doing nothing would in fact be the best option since both

Table 8.1 Scenario 1: Life Years Resulting from Two Alternative Policies

	Policy effects in life expectancies		
(1) Number of persons	(2) *Do Nothing*	(3) Policy *A*	(4) Policy *B*
20	10	10	11
1	10	30	10

Numbers in columns 2-4 represent life expectancies.

utilitarianism with a zero social discount rate neither policy would be preferred, prioritarianism would select policy *A*, while utilitarianism with a positive social discount rate would choose policy *B*. I believe, in this scenario, policy *A* to be the better policy particularly given the previously presented arguments against a positive social discount rate. Indeed, considering solely gains to longevity, there are many cases where a positive social discount rate would lead to us preferring an outcome with smaller absolute gains. In any case, this example highlights how these factors make a difference to the policies we ultimately select.

Under the Priority View, neonates, in general, would be given high weights owing to the long length of life and QALY gains that result from mortality and morbidity reductions in this group. However, this is not necessarily an argument to support policies that reduce the mortality rate (as many of the studies in thesis have focussed)—those at the margin for risk of mortality are often among the sickest infants, reduction of the mortality rate further may serve to increase health inequalities among newborns, and in later life. This may seem to undermine the Priority View. However, this is the reason why whole life welfare should be taken into account, an infant at the margin for risk of mortality may not be the 'worst off' if that infant will only go on to live a short period in poor health (and no other policy or technology exists that could improve this). The worst-off infant may be the one at lower risk of mortality but for whom the consequences of death would be a much greater loss (for example, in scenario 1).

An alternative formulation of egalitarianism is known as luck egalitarianism. Under a luck egalitarian philosophy, it would be bad if some people were worse off than

policies *A* and *B* increase inequality despite the fact that they would be worse for no-one. I disregard such arguments here for the sake of brevity. Parfit (1997) considers such positions.

others, except in the case of those inequalities arising through voluntary choice of faulty conduct (Anderson, 1999). Luck egalitarianism has been used to argue for the giving of less priority to those who bear some responsibility for their condition in decisions regarding the distribution of healthcare resources, such as smokers suffering from smoking related illnesses or the obese. There are numerous objections to luck egalitarianism which are beyond the scope of this discussion (see, for example, Anderson (1999) and Lippert-Rasmussen (2001)). Needless to say, given the lack of agency on the part of a newborn infant this group should be weighted among the most highly under luck egalitarianism.

Finally, democratic equality sees justice as equality by providing each individual the freedom to participate in democratic self-government. This concept is similar to other philosophies which promote equality of opportunity as well as the Capability Approach, which focusses on the capability of individuals to lead the lives that they have reason to value (Sen, 1985). These various formulations differ somewhat in their approaches and how they would be operationalised but all have the characteristic that they see the distribution of resources as only of instrumental importance, and that each person should be at a basic threshold level in order for them to enjoy certain freedoms or opportunities. Health is often seen as of more fundamental importance than wealth so that society is more averse to inequalities in health than in wealth (Anand, 2004; Sen, 1985). The goal of health policy may then be to maintain a certain level of health in the population. However, maintenance of this basic level may come at too high a cost and we may forego large gains to population health elsewhere in pursuit of this threshold. For this reason, the Priority View may be preferred, since it takes this into account (Parfit, 1997).

This section has discussed the various arguments about distributional justice, and while no definitive arguments are presented here, I would defend the Priority View. However, what is highlighted here, is that benefits accruing to neonates and infants may be weighted relatively highly compared to a purely efficiency based system. In part, this is due to the duration of the outcomes from early life interventions and how even small changes may be amplified over the course of a lifetime. The characteris-

tics of a person at birth have been previously shown to affect educational, health, and labour market outcomes in later life (Black et al., 2007). The Fetal Origins Hypothesis posits that the nine months *in utero* are one of the most critical periods in shaping a person's future health trajectory (Almond and Currie, 2011) which may, in part, explain the persistence of intergenerational health inequalities among other factors Marmot (2004). In addition, Chapter 7 showed that increases to the local unemployment rate increases the proportion of live births admitted to neonatal healthcare. Other studies have also shown that maternal nutrition or paternal job losses also led to deteriorations in infant health at birth (Almond and Mazumder, 2011; Lindo, 2011). Given the implied relationships between health, labour market outcomes, and subsequent infant health at birth along with the demonstrable persistence of intergenerational health inequalities, this may suggest that the welfare losses due to, for example, reductions in household income, may be much larger than analyses typically estimate. The magnitude of the (value of the) effect depends on the intergenerational social discount rate. In fact, it may be argued, that once intergenerational welfare changes are incorporated into analyses, the cost-effectiveness of various technologies and policies may be altered. Indeed, it may even suggest that, once intergenerational welfare effects are taken into account, selecting the policies with the largest welfare gains may also improve overall health equalities (by the same principle that the Priority View increases equality).

8.4 Centralisation of Neonatal Healthcare Services

As an example of how the preceding discussion may affect decision making with regards to neonatal healthcare policies, let us consider the policy of centralisation. Chapter 4 demonstrated that very preterm infants admitted to a high volume neonatal unit at the hospital of birth were at lower risk of mortality that their counterparts admitted to a low volume neonatal unit at the hospital of birth. This effect appeared to be driven by the reduction in the risk of mortality for extremely preterm infants rather infants born at a later gestation. No effects were observed for the considered mor-

bidities. On this basis of this evidence, a centralisation policy may be advocated, as authors in, for example, the United States have done in response to similar evidence observed there (Phibbs, 2012; Phibbs et al., 2007). Let us assume that the centralisation of neonatal healthcare would bear no direct cost to the healthcare system. The evidence from Chapter 4 suggests that centralisation of neonatal healthcare in England would only benefit extremely preterm infants; extremely preterm infants (defined according to Chapter 4 as those born at $\leq 26^{+6}$ weeks gestation) only make up 2.1% of neonatal admissions but comprise 40.9% of all deaths on neonatal units (see also table 3 in Chapter 2). Table 10 in Chapter 4 shows that 46.4% of all very preterm infants were born in high volume neonatal units, of which 5.5% died. Assuming, under a centralisation policy, all very preterm infants would be born in and admitted to high volume neonatal units at the hospital of birth, then the results in Chapter 4 suggest centralisation would prevent approximately 58 deaths per year.

The effects of a centralisation policy differ from those in Table 8.1; centralisation would provide a potentially large gain for some infants for a potentially small cost borne by a large number of infants. Centralisation could be seen as akin to the following scenario, where, instead of life expectancy we consider a measure of whole life welfare (for example, a number of *Utils*, without loss of generalilty, where 100 is equivalent to a full life in good health):

Table 8.2 Scenario 2: Effects of a Centralisation Policy

	Utility Resulting from Policies	
(1) Number of persons	(2) *Do nothing*	(3) *Centralisation*
20	95	*N*
1	1	80

Numbers represent undiscounted lifetime number of *Utils* (or other measure of welfare).

In this scenario,[7] we must decide whether to choose the policy of centralisation where the majority of infants have a small welfare loss (*N* being less than 95) and one infant has a large gain. For many persons, if *N* is in the range of 85 to 94, I

[7]These are exactly the sorts of scenarios considered by Parfit (1997).

believe we would have difficulty choosing between the two policies. Scenario 2 is meant to represent the choice of centralisation policies faced by a decision maker. As I previously argued, we should consider whole life welfare rather than short term welfare. In this case, N is likely to be very close to 95, provided there are no negative health consequences of centralisation to healthier infants. Indeed, it is possible that the main cost to these infants is borne by their families in travelling further to receive neonatal care which, over the course of a life may become a very small effect, even with a positive social discount rate.

It must also be considered that the welfare gains to the infants that benefit from centralisation may not be particularly large, in view of a whole life. The disability free survival rate for infants born at less that 27 weeks gestation is only 41% for those who survive to discharge (Costeloe et al., 2012). The resulting quality of life among those infants is likely to be (significantly) less than the general population.[8] These infants are also likely to generate high costs to the state in terms of further healthcare requirements, educational needs, and community care (Mangham et al., 2009). Chapter 5 shows though, that even after taking into account these extra costs, the costs per life year gained at the margin for neonatal healthcare may be smaller than the ICERs estimated for other programmes of healthcare (Claxton et al., 2013).

In any case, if we take the Priority View, we ought to weight benefiting the worse-off infants higher than those among the better-off infants. Furthermore, taking a whole life view, the costs incurred among better-off infants due to centralisation are likely to be relatively small. Moreover, the state could facilitate parents by providing support to parents for increased travel distances associated with centralisation so that N in Scenario 2 is as close to 95 as possible.

In conclusion, I would argue in support of centralisation given the increased weight of the claims to healthcare that the worst-off infants may have, however, it is essential to establish whole life outcomes as accurately as possible for these infants, the current paucity of data in this regard makes evaluation of such a policy difficult. In any case,

[8]However, individuals with disabilities often adapt to their circumstances and experience a quality of life on par with the rest of the population, this is known as hedonic adaptation (Frederick and Lowenstein, 2003).

the primary point of this chapter is to provide a framework to inform the overall policy conclusions that can be made from the evidence presented in this thesis and in so doing identify the areas required for future research to strengthen the ensuing claims made. The following, and final, chapter, draws together the evidence provided in this thesis and provides a number of policy conclusions and recommendations for future research.

Chapter 9

Conclusions

The aim of this thesis was to explore the economic and healthcare related determinants of infant health at birth and in so doing provide evidence that may be utilised by policy makers to improve the health of infants treated on neonatal units. The previous literature on this subject is relatively rich and there has generally been a consensus over the effects on infant clinical outcomes of neonatal unit characteristics such as unit volume, measured in terms of patient or procedure numbers, and nurse to patient ratios. However, since 2003, neonatal units in England have been organised into networks in order to improve access to high volume and designation centres. None of the previous research has been conducted in such a setting, including work from the UK, of which the most recent used data dating from from 1998-9. This thesis also presented some of the first analyses to make use of the National Neonatal Research Database (NNRD), a novel and rich source of clinical data. Previous studies have been limited in terms of the patient population or neonatal units covered by the available data.

This thesis sought to answer four specific questions relating to the health of newborn infants:

- What are the effects of neonatal unit designation and volume on the clinical outcomes of infants treated within those units?
- What are the returns to neonatal healthcare expenditure currently, in terms of health outcomes, being achieved within neonatal healthcare in England?

- What is the effect of the nurse to patient ratio on the risk of mortality among infants receiving neonatal intensive care?
- Do local economic conditions at the time of conception affect the health of infants at birth?

The NNRD was utilised in the empirical chapters of this thesis to attempt to provide answers to these questions. The methods and data utilised here provide these analyses with strengths over the previous literature in this area. For example, Chapter 5 estimated the health effects of changes to neonatal healthcare expenditure. Previous studies to examine this question have employed aggregate, local area level data, which I argued is not necessarily suitable to identify the effect of interest. Chapter 5 showed how routinely collected, individual level patient data, such as that in the NNRD, along with other publicly available datasets may be used for enquiries of this nature. However, it must also be recognised that there are limitations to the analyses in this thesis and that these may limit the generalisability of the results, the specificity of the policy conclusions that can be made from the results, or at least alter the interpretation of the results in some manner—these issues will be discussed in this concluding chapter. Nonetheless, this research has important policy implications, and suggests many possible future avenues of research.

One of the greatest changes to the structure of neonatal healthcare in England in recent years was the introduction of managed clinical networks (MCN) in 2003. A networked approach to neonatal care was adopted, in part, in response to evidence that very low birth weight and very preterm infants born in units that treated greater numbers of patients or which provided higher intensity care were at a lower risk of mortality than their counterparts born elsewhere (Cifuentes et al., 2002; Department of Health, 2003). It was decided that centralisation of neonatal healthcare services, by closing smaller or lower intensity units, would reduce equity of access to neonatal healthcare. However, since the formation of MCNs, now called Operational Delivery Networks, there has been no research to determine whether the advantage of a high volume neonatal unit at the place of birth has been eliminated or reduced by MCNs.

This was the question that Chapter 4 sought to answer. It was argued in Chapter 1, that MCNs would be unlikely to be able to replicate the benefits of a centralised system partly due to the importance of the first few hours of an infant's life. Indeed, the first sixty minutes are often referred to as the Golden Hour in which a number of complex, team-oriented, task-based activities occur, the success of which are vital in determining an infant's chances of survival (Doyle and Bradshaw, 2012). Unit volume and designation may affect the quality of the care provided during this, and subsequent, periods of infant care. Thus, the place of birth of an infant may be of great importance for which postnatal transfers may not be a suitable replacement.

There are two possible mechanisms by which neonatal unit volume may have a causal effect on clinical outcomes: economies of scale and learning by doing (Luft et al., 1979, 1987). The former mechanism means that the long run average cost of neonatal care provision is lower for higher volume neonatal units so that they can afford greater inputs to neonatal care. If economies of scale were the only mechanism acting, then increases in expenditure on neonatal healthcare in smaller neonatal units would eliminate the advantages of a larger neonatal unit at the hospital of birth (without necessarily being a cost-effective way of improving infant clinical outcomes). The topic of Chapter 5 is the effect of expenditure on neonatal healthcare on the risk of mortality of infants admitted to the neonatal unit at the hospital of birth. This chapter obtained estimates of the cost per care day at each level of care at a large number of NHS Trusts from the NHS Reference Costs database. These were matched to the individual level data in the NNRD. It was shown that infants admitted to a neonatal unit at the hospital of birth with higher total expenditure, and thus greater labour and capital inputs to healthcare production, were at a lower risk of mortality than their counterparts admitted to lower spending units. It was found that this effect differed by both patient group and neonatal unit volume.

One of the ways in which increased neonatal healthcare expenditure may translate into reduced risk of mortality among neonates is through increased labour inputs to neonatal units. One recent study found that 54% of shifts on neonatal units in a particular MCN were understaffed with respect to the guidelines provided by the British

Association of Perinatal Medicine (BAPM) (British Association of Perinatal Medicine, 2010). This has led some organisations to campaign for greater funding for nursing labour on neonatal units (Bliss, 2011). Furthermore, given that cots on neonatal units are often closed to new admissions due to a lack of available labour to staff these cots, rather than there being too many staff for each cot (Parmanum et al., 2000), this may suggest that neonatal units are constrained in their supply of labour rather than capital inputs. For infants receiving intensive care in England, BAPM recommend a one to one nurse to patient ratio, however, the NNRD data reveal that only 10% of IC care days have one to one nursing provided. The aim of Chapter 6 was to determine the effect of one to one nurse provision on the outcomes of neonates receiving intensive care on neonatal units. The results in this chapter showed that a ten percentage point increase in the proportion of intensive care days receiving one to one nursing leads to a reduction in the mortality rate of 0.56 percentage points. This concurs with the findings of a number of other studies that investigated the effect of nurse to patient ratios on clinical outcomes in neonatal units (Sherenian et al., 2013).

This thesis has presented evidence to show that increased inputs to neonatal care may reduce the risk of mortality among infants admitted to neonatal units. This may explain the advantage of high volume neonatal units. However, if learning by doing plays a role in mediating the causal effect of volume on health outcomes, then the strategy of increasing inputs to care in smaller units will only be partially successful. Learning by doing refers to the increase in specific human capital among the workforce of a high volume neonatal unit that comes through increased experience of treating infants at high risk of mortality. This human capital is likely to degrade over time if it is unused (Gaynor et al., 2005; Sfekas, 2009), such as if a physician moved to a smaller neonatal unit, and is therefore not likely to be transferable between units. Moreover, it is also likely to be inimitable and non-substitutable, since the training received by neonatologists and nursing staff is the same in all units. Chapter 4 found that infants admitted to high volume neonatal units at the hospital of birth were at a lower risk of mortality than their counterparts admitted to low volume neonatal units. In particular, very preterm infants (born at less than 33 weeks gestation) were 32% less

likely to die if admitted to a high volume neonatal at the hospital of birth as opposed to a low volume neonatal unit. Importantly, when the sample of very preterm infants was separated into those born at <27 weeks gestation and those born at 27-32 weeks gestation, a statistically significant effect was only observed in the former group. It was found that unit designation did not have a *ceteris parabis* causal effect on infant clinical outcomes.

These results lend support to the preceding arguments in this chapter but are not able to differentiate between the economies of scale hypothesis and the learning by doing hypothesis. Under increasing returns to scale, an increase in expenditure should correspond to a greater increase in inputs in a high volume neonatal unit than in a lower volume neonatal unit, and thus a greater reduction the risk of mortality. In Chapter 5 it was shown that the effects of a *ceteris parabis* increase in neonatal expenditure of the risk of mortality for infants admitted to that unit were greater in high volume units than low volume units. This demonstrates that high volume neonatal units may be more technically efficient than the low volume counterparts, but whether this is driven by economies of scale or learning by doing remains an important topic for future research. In either case, this may suggest that the benefits of centralisation may not be replicated by increased funding to lower volume neonatal units.

The evidence presented in this thesis is of particular importance today. As was evidenced in Chapter 7, the proportion of live births admitted to neonatal specialist care increased between 2007 and 2011 from 7.4% of births in 2007 to 9.0% of births in 2011 while the birth rate did not change over the same period. And, as that chapter showed, while the increase in admissions was generally of babies born closer to term gestation or normal birth weight, the proportion of live births admitted to neonatal care and receiving intensive care increased as well. If the number of admissions to neonatal healthcare increase but the number of nursing staff do not, then this may have deleterious consequences on the clinical outcomes of admissions to neonatal units as the staff to patient ratios would decline.

The aim of Chapter 7 was to examine the relationship between local economic conditions at conception and infant health at birth, which was measured by the rate

of admission to neonatal units among live births. This is particularly important here for the planning of neonatal healthcare services; if it is expected that admissions to neonatal units may increase, then this would provide additional evidence to increase funding for labour inputs to neonatal healthcare. Tucker (2002) found an association between neonatal unit occupancy and risk of mortality which further suggests that neonatal units need to be able to expand capacity in response to changes in the requirement for their care. Using the local unemployment rate at the time of conception as a measure of local economic conditions, Chapter 7 found that a one percentage point increase in the unemployment rate would increase the admissions rate among live births by 0.36 percentage points. Moreover, extensions to this analysis only found evidence of this effect in the most socio-economically deprived areas. It was argued that this was a causal effect. On this basis, it may be simply predicted that were there to be a fall in the unemployment rate, then the proportion of live births admitted to neonatal care would decrease. However, given the complex nature of the relationship between socio-economic conditions and population health, specific predictions regarding exact numbers of admissions are generally not possible to make.

A wide variety of findings are presented in this thesis. Although, translating them into implementable policies requires consideration of a number of key normative issues. Chapter 8 surveys these issues and provided some simple examples to illustrate the difficulty with which the results of this thesis can be effectively used in the formulation of policy. However, it is in identifying these difficulties that the recommendations for future research can be identified. The results of Chapter 4 certainly provide evidence that suggests that a policy that leads to an increase in the proportion of extremely preterm births that occur in hospitals with high volume neonatal units would reduce the risk of mortality among this group. Such a policy may involve increased provision of *in utero* inter-hospital transfers or the centralisation of neonatal healthcare. However, whether such a policy ought to be enacted requires consideration of its long term consequences. Indeed, after some consideration, I argued in Chapter 8 that the evidence of this thesis tentatively supported a centralisation policy, but that what was required was better data on the long term outcomes of the infants that would be affected by

such a policy. Indeed, one of the weaknesses of this study, is the lack of follow-up data for infants that appear in the NNRD. However, this is already changing—ongoing work aims to link NNRD patient records to subsequent admissions records in other hospital units, presenting an important avenue of future research that may strengthen the evidence presented here.

Further policy recommendations follow from the results of this thesis. Chapter 6 lends support to the BAPM guideline of one to one nursing for infants receiving neonatal intensive care in England. However, it is not known whether this is the *optimal* ratio for these patients, nor is it known how re-allocating nursing staff from other tasks to one to one nursing would affect the clinical outcomes of other patients. Since nursing shifts are often understaffed on neonatal units (Pillay et al., 2011), and only a small number of intensive care care days meet the BAPM guideline, the evidence of Chapter 6 lends support to policies aimed at increasing the provision of nursing labour on neonatal units. Nevertheless, to better determine optimal nurse to patient ratios and calculate the cost-effectiveness of increasing nurse labour provision, data on the numbers of staff over different points in time in multiple units would be required. This is an important area for future research in this field.

Chapter 7 also provides evidence to suggest that interventions targeted at households that aim to improve the health of newborn infants may be more effective if targeted at certain areas. In particular, it was shown that increases to the unemployment rate led to increases in the proportion of live births admitted to neonatal healthcare only in the most deprived areas. Thus, if we wish to develop policies to reduce the burden on neonatal units, by improving the health of infants at birth, these may be better targeted at the most socio-economically deprived. It was shown that it was increases to the unemployment rate among economically active males that drove the observed effect in this chapter. Again, identifying the reasons for this is an important area of future research in order to guide policy. One potential explanation could be the deterioration in maternal nutrition due to a reduction in household income, which has been shown to impact infant health at birth (Almond and Mazumder, 2011), which might suggest a policy of improved unemployment benefit for households or the provision of

nutrition to expectant mothers, for example.

The analyses in this thesis have a number of strengths over previous studies in this area. As previously discussed, one of the key strengths of this thesis is the use of the NNRD, which represents the vast majority of admissions to neonatal care in the country. Previous studies have had use of rich data, for example, Phibbs et al. (2007) examined all very low birth weight births in California over a nine year period by linking birth and death certificates with hospital discharge abstracts. While these data are rich and provide a valuable resource; the NNRD enables analysis of the entire patient population and include data, such as day to day information on the level of care provided as well as a multitude of outcomes that other studies have not had access to. Moreover, the NNRD data is collected almost in real time, permitting up to the moment analysis. The most recent data used by Phibbs et al. (2007) was collected in 2000, seven years prior to publication. Over the same seven year period, neonatal care in England was reorganised dramatically, highlighting the need for up to date data. Consider that since 2013, the organisation and management of neonatal networks has again changed to some extent following the Health and Social Care Act (2013) perhaps warranting further analysis. The richness of the data in the NNRD also permitted the analysis of one to one nursing for neonatal intensive care as daily one to one nursing ratios were recorded. No previous study has been able to do this.

The NNRD, however, is not perfect and weaknesses arising from its use must be acknowledged. The NNRD only contains data on those infants admitted to neonatal units in the UK. Inferences arising from the results of these analyses can therefore only be made with respect to this patient population. For example, in Chapter 4 it was shown that very preterm infants admitted to a high volume neonatal unit at the hospital of birth were at lower risk of any in hospital mortality than their counterparts admitted to low volume neonatal units at the hospital of birth, inferences cannot be made about the effects of birth in a hospital with a high volume neonatal unit for all very preterm infants—those infants who died in hospital prior to admission are not observed in the NNRD. Nonetheless, high volume neonatal units are often collocated with high volume delivery units, infants born on which have been shown be at lower

risk of mortality than infants born on low volume delivery units (Heller et al., 2002; Moster et al., 1999). This may suggest that the results of Chapter 4 underestimate the benefit of birth in a hospital with a high volume neonatal unit. Similar criticisms may be applied to chapters 5 and 6. Hospitals with higher nurse to patient ratios on neonatal units may or may not have higher nurse to patient ratios on delivery units which may affect mortality on delivery units and hence the sample of infants that are eventually admitted onto neonatal units. Analysis of the effects of neonatal unit characteristics in the context of the wider level of resourcing in both the hospital and wider healthcare environment is thus an essential avenue for future research.

There are a number of missing data in the NNRD (see Table 2.2). This may present an issue for the empirical analyses, particularly if the mechanism that generates the missing data is related to some variable that is not included in the statistical models.[1] Fortunately, the proportions of data that are missing for many of the key variables that feature in models of neonatal clinical outcomes, such as gestational age and birth weight (Medlock et al., 2011), are very low. Where missing data may potentially lead to inconsistency of the estimators used, different techniques from the econometric literature are employed; in Chapter 7, it was shown that there may be systematic differences between the areas near neonatal units that contribute their patient data to the NNRD and those that do not. A correction for this selection problem was employed (Wooldridge, 1995). Further development of methods to potentially address the problems caused by missing data in the NNRD are thus important to strengthen the results presented here. Even so, the rate of missing data has reduced over time, being indicative of an overall improvement over time in data collection for the NNRD.

The NNRD represents a rich source of data for future research in this area. Continued improvement of data quality and reduction of missing data along with linkages to other healthcare and demographic data sources will only serve to improve future research. Caution must be advised, however, about the use of such large data sources as these. Increased numbers of observations typically result in improved precision of es-

[1] That is the data are missing not at random (MNAR)—the distinctions between different types of missingness are given in Section 2.2.1 in Chapter 2.

timates of parameters in statistical models, yet this also means that we are more likely to detect smaller, potentially clinically insignificant results.

The methodology utilised in the various empirical chapters also represents a strength of the thesis. Chapter 3 reviewed the methodology of the previous studies in this area and found that only a small minority of the surveyed studies took into account the possible endogeneity of unit level variables (for example, Lorch et al. (2012)). Patients are frequently transferred between units on the basis of the ability of those units to provide the care appropriate for those patients. Given that we cannot observed patient health perfectly, this leads to the aforementioned endogeneity. The methods used in this thesis have taken this into account and therefore the results presented are arguably causal effects, which it is essential to identify if accurate policy recommendations are to be made. I exploited the strong, exogenous preference of individuals to go to their nearest hospital for healthcare, and utilised instrumental variables approaches to account for unobserved patient heterogeneity. However, in doing so, the interpretation of the estimated effects must be considered. The effects here only apply to those individuals or units whose treatment status is altered by the instrumental variable utilised—these are the compliers with the instrument. Chapter 4 argued that this represented the vast majority of the patient population. Nonetheless, sweeping generalisation of these results should be avoided. Chapter 5 also demonstrated how patient level data, such as the NNRD, may be utilised to address an important research aim in health economics—namely the estimation of the cost-effectiveness threshold for economic evaluation of healthcare technologies. This is the first such study to do so.

A number of future research aims have already been identified in this chapter. It was discussed that linkage of the NNRD data with other healthcare and demographic data sources will expand the range of questions that can be investigated with these data. In particular, neonatal units operate in conjunction with other units in the hospital, incorporating the effects of these units into analyses is important. As Chapter 8 discussed, the most appropriate metric for the evaluation of neonatal unit interventions are whole life outcomes taking into account quality of life. Data regarding long term outcomes are often unavailable to researchers. The data collected from infants

who have subsequent contact with the healthcare system post-discharge from a neonatal may provide important insights into the longer term effects of neonatal healthcare interventions. Increasing the labour inputs to neonatal units may have an important role to play in improving neonatal clinical outcomes, this was shown to some extent in Chapter 6; further analysis and more detailed data on neonatal unit staffing are required to be able to determine the optimal level of staffing. Understanding the complex pathways that lead to poor infant health at birth is also crucial to the improvement of population health in the long run. In all these cases, the potential benefits of policies that improve infant health may be large given the whole life duration over which these policies may act, emphasizing the importance of this research.

Neonatal healthcare has made great advances in the last few decades when the mortality rate is considered. This thesis provides evidence to suggest that further advances can yet be made. Many of the future improvements to neonatal clinical outcomes in England may come due to improvements in the organisation, planning, and funding of neonatal care. In any case, further data is required on long term outcomes of the recipients of neonatal healthcare to truly weigh up alternative policies. As such, the future gains to neonatal healthcare may not be in reducing the mortality rate further, but improving the quality of life of those that receive neonatal healthcare.

References

Abdel-Latif, M. E., Bajuk, B., Oei, J., and Lui, K. (2006). Mortality and morbidities among very premature infants admitted after hours in an Australian neonatal intensive care unit network. *Pediatrics*, 117(5):1632–9.

Allen, M. C., Donohue, P. K., and Dusman, A. E. (1993). The limit of viability–neonatal outcome of infants born at 22 to 25 weeks' gestation. *The New England journal of medicine*, 329(22):1597–601.

Allison, P. D. (2001). *Missing Data*. SAGE Publications, Inc, 1st editio edition.

Almond, D. and Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3):153–172.

Almond, D., Doyle, J. J., Kowalski, a. E., and Williams, H. (2011). The Role of Hospital Heterogeneity in Measuring Marginal Returns to Medical Care: A Reply to Barreca, Guldi, Lindo, and Waddell. *The Quarterly Journal of Economics*, 126(4):2125–2131.

Almond, D., Doyle, Jr., J. J., Kowalski, A. E., and Williams, H. (2010). Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns *. *Quarterly Journal of Economics*, 125(2):591–634.

Almond, D. and Mazumder, B. (2011). Health Capital and the Prenatal Environment: The Effect of Ramadan Observance During Pregnancy. *American Economic Journal: Applied Economics*, 3(4):56–85.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1):151–184.

American Academy of Pediatrics Committee on Fetus and Newborn and American College of Obstetricians and Gynecologists Committee on Obstetric Practice (2012). *Guidelines for Perinatal Care*. 7th editio edition.

American College of Obstetricians and Gynaecologists (2013). Committee Opinion No 579: Definition of Term Pregnancy. *Obstetrics and gynecology*, 122(5):1139–1140.

Anand, S. (2004). The Concern for Equity in Health. In Anand, S., Peter, F., and Sen, A., editors, *Public Health, Ethics, and Equity*, chapter 1, pages 15–20. Oxford University Press, New York, 1st editio edition.

Anderson, E. S. (1999). What Is the Point of Equality?

Appleby, J. (2012). Rises in healthcare spending: where will it end? *BMJ*, 345(nov01 1):e7127–e7127.

Arrow, K. (1962). The Economic Implications of Learning by Doing. *American Economic Review*, 29:155–173.

Baba, K., Shibata, R., and Sibuya, M. (2004). PARTIAL CORRELATION AND CONDITIONAL CORRELATION AS MEASURES OF CONDITIONAL INDEPENDENCE. *Australian & New Zealand Journal of Statistics*, 46(4):657–664.

Baiocchi, M., Small, D. S., Lorch, S., and Rosenbaum, P. R. (2010). Building a Stronger Instrument in an Observational Study of Perinatal Care for Premature Infants. *Journal of the American Statistical Association*, 105(492):1285–1296.

Baker, L. C. and Phibbs, C. S. (2002). Managed care, technology adoption, and health care: the adoption of neonatal intensive care. *The Rand journal of economics*, 33(3):524–48.

Barker, D., Rosenthal, G., and Cram, P. (2011). Simultaneous relationships between procedure volume and mortality: do they bias studies of mortality at specialty hospitals? *Health economics*, 20(5):505–18.

Barnsley, P., Towse, A., Karlsberg Schaffer, S., and Sussex, J. (2013). Critique of the CHE Research Paper on NICE Thresholds.

Barreca, a. I., Guldi, M., Lindo, J. M., and Waddell, G. R. (2011). Saving Babies? Revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, 126(4):2117–2123.

Bartels, D. B., Wypij, D., Wenzlaff, P., Dammann, O., and Poets, C. F. (2006). Hospital volume and neonatal mortality among very low birth weight infants. *Pediatrics*, 117(6):2206–14.

Becker, G. S., Philipson, T. J., and Soares, R. R. (2005). The Quantity and Quality of Life and the Evolution of World Inequality. *American Economic Review*, 95(1):277–291.

Behrman, R. E. (1971). Fetal and Neonatal Mortality in White Middle Class Infants. *American Journal of Diseases of Children*, 121(6):486.

Bell, E. F., Hansen, N. I., Morriss, F. H., Stoll, B. J., Ambalavanan, N., Gould, J. B., Laptook, A. R., Walsh, M. C., Carlo, W. a., Shankaran, S., Das, A., and Higgins, R. D. (2010). Impact of timing of birth and resident duty-hour restrictions on outcomes for small preterm infants. *Pediatrics*, 126(2):222–31.

Bharadwaj, P., Lø ken, K. V., and Neilson, C. (2013). Early Life Health Interventions and Academic Achievement. *American Economic Review*, 103(5):1862–1891.

Binder, S., Hill, K., Meinzen-Derr, J., Greenberg, J. M., and Narendran, V. (2011). Increasing VLBW deliveries at subspecialty perinatal centers via perinatal outreach. *Pediatrics*, 127(3):487–93.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007). From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *The Quarterly Journal of Economics*, 122(1):409–439.

Bliss (2011). SOS: Save our special care babies, save our specialist nurses. Technical report, Bliss, London.

Bobinac, A., van Exel, N. J. A., Rutten, F. F. H., and Brouwer, W. B. F. (2012). Inquiry into the relationship between equity weights and the value of the QALY. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 15(8):1119–26.

Bode, M. M., O'shea, T. M., Metzguer, K. R., and Stiles, a. D. (2001). Perinatal regionalization and neonatal mortality in North Carolina, 1968-1994. *American journal of obstetrics and gynecology*, 184(6):1302–7.

Boss, R. D., Kinsman, H. I., and Donohue, P. K. (2012). Health-related quality of life for infants in the neonatal intensive care unit. *Journal of perinatology : official journal of the California Perinatal Association*, 32(12):901–6.

Bowles, S. (1998). Endogenous Preferences : The Cultural Consequences of Markets and other Economic Institutions. *Journal of Economic Literature*, 36:75–111.

Bowles, S. and Polanía-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50:368–425.

Boxwell, G., editor (2010). *Neonatal Intensive Care Nursing*. Routledge, Abingdon, Oxon, 2nd editio edition.

Boyle, E. M., Poulsen, G., Field, D. J., Kurinczuk, J. J., Wolke, D., Alfirevic, Z., and Quigley, M. A. (2012). Effects of gestational age at birth on health outcomes at 3 and 5 years of age: population based cohort study. *BMJ (Clinical research ed.)*, 344:e896.

British Association of Perinatal Medicine (2010). Standards for hospitals providing neonatal intensive and high dependency care (3rd Edition). Technical report, BAPM, London.

British Association of Perinatal Medicine (2011). Categories of Care 2011.

Bunker, J. P. (1995). Medicine matters after all. *Journal of the Royal College of Physicians of London*, 29(2):105–12.

Bunker, J. P. (2001). The role of medical care in contributing to health improvements within societies. *International journal of epidemiology*, 30(6):1260–3.

Bunker, J. P., Frazier, H. S., and Mosteller, F. (1994). Improving health: measuring effects of medical care. *The Milbank quarterly*, 72(2):225–58.

Callaghan, L. a., Cartwright, D. W., O'Rourke, P., and Davies, M. W. (2003). Infant to staff ratios and risk of mortality in very low birthweight infants. *Archives of disease in childhood. Fetal and neonatal edition*, 88(2):F94–7.

Cameron, A. C. and Trivedi, P. (2005a). 21.6 Fixed Effects Model. In *Microeconometrics: Methods and Applications*, chapter 21, pages 726–733. Cambridge University Press, 1st editio edition.

Cameron, A. C. and Trivedi, P. (2005b). 22.2 GMM Estimation of Linear Panel Models. In *Microeconometrics: Methods and Applications*. Cambridge University Press, 1st editio edition.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In Christofides, L., Grant, E., and Swidinsky, R., editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. University of Toronto Press, Toronto, 1st editio edition.

Carlo, W. (2011). Prematurity and intrauterine growth restriction. In Kliegman, R., Behrman, R. E., Jenson, H., and Stanton, B., editors, *Nelson Textbook of Pediatrics*, chapter 91.2. Saunders Elsevier, Philadelphia, PA, 19th ed. edition.

Casey, B. M., McIntire, D. D., and Leveno, K. J. (2001). The continuing value of the Apgar score for the assessment of newborn infants. *The New England journal of medicine*, 344(7):467–471.

Chang, C.-L. and McAleer, M. (2013). Ranking journal quality by harmonic mean of ranks: an application to ISI statistics & probability. *Statistica Neerlandica*, 67(1):27–53.

Chung, J. H., Phibbs, C. S., Boscardin, W. J., Kominski, G. F., Ortega, a. N., Gregory, K. D., and Needleman, J. (2011). Examining the effect of hospital-level factors on mortality of very low birth weight infants using multilevel modeling. *Journal of perinatology : official journal of the California Perinatal Association*, 31(12):770–5.

Chung, J. H., Phibbs, C. S., Boscardin, W. J., Kominski, G. F., Ortega, A. N., and Needleman, J. (2010). The effect of neonatal intensive care level and hospital volume on mortality of very low birth weight infants. *Medical care*, 48(7):635–44.

Chung, M.-Y., Fang, P.-C., Chung, C.-H., Chen, C.-C., Hwang, K.-P., and Chen, F.-S. (2009). Comparison of neonatal outcome for inborn and outborn very low-birthweight preterm infants. *Pediatrics international : official journal of the Japan Pediatric Society*, 51(2):233–6.

Cifuentes, J., Bronstein, J., Phibbs, C. S., Phibbs, R. H., Schmitt, S. K., and Carlo, W. a. (2002). Mortality in Low Birth Weight Infants According to Level of Neonatal Care at Hospital of Birth. *Pediatrics*, 109(5):745–751.

Cimiotti, J. P., Haas, J., Saiman, L., and Larson, E. L. (2006a). Impact of staffing on bloodstream infections in the neonatal intensive care unit. *Archives of pediatrics & adolescent medicine*, 160(8):832–6.

Cimiotti, J. P., Haas, J., Saiman, L., and Larson, E. L. (2006b). Impact of staffing on bloodstream infections in the neonatal intensive care unit. *Archives of pediatrics & adolescent medicine*, 160(8):832–6.

Clark, S. L., Frye, D. R., Meyers, J. A., Belfort, M. A., Dildy, G. A., Kofford, S., Englebright, J., and Perlin, J. A. (2010). Reduction in elective delivery at <39 weeks of gestation: comparative effectiveness of 3 approaches to change and the impact on neonatal intensive care admission and stillbirth. *American journal of obstetrics and gynecology*, 203(5):449.e1–6.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P. C., and Sculpher, M. (2013). Methods for estimation of the NICE cost-effectiveness threshold.

Claxton, K., Paulden, M., Gravelle, H., Brouwer, W., and Culyer, A. J. (2011). Discounting and decision making in the economic evaluation of health-care technologies. *Health Economics*, 20:2–15.

Cole, T. J., Hey, E., and Richmond, S. (2010). The PREM score: a graphical tool for predicting survival in very preterm births. *Archives of disease in childhood. Fetal and neonatal edition*, 95(1):F14–9.

Costa, D. L. and Kahn, M. E. (2004). Changes in the value of life, 1940-1980. *Journal of Risk and Uncertainty*, 29(2):159–180.

Costeloe, K. L., Hennessy, E. M., Haider, S., Stacey, F., Marlow, N., and Draper, E. S. (2012). Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ (Clinical research ed.)*, 345:e7976.

Cowen, T. (1990). Distribution in fixed and variable number problems. *Social Choice and Welfare*, 7(1):47–56.

Cowen, T. (2001). What is the Correct Intergenerational Discount Rate?

Cowen, T. and Parfit, D. (1992). Against the Social Discount Rate. In Laslett, P. and Fishkin, J., editors, *Philosophy, Politics, and Society*, pages 144–161. Yale University Press, New Haven, sixth edit edition.

Culyer, A. J. (2010). Entry for Quality Adjusted Life Year. In *The Dictionary of Health Economics*, pages 428–9. Edward Elgar Publishing, London, 2nd editio edition.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.

Currie, J., Stablie, M., Manivong, P., and Roos, L. L. (2010). Child Health and Young Adult Outcomes. *Journal of Human Resources*, 45:517–548.

Cutler, D. M., Lleras-Muney, A., and Vogi, T. (2011). Socioeconomic status and health: Dimensions and mechanisms. In Glied, S. and Smith, P. C., editors, *The Oxford Handbook of Health Economics*, chapter 7, pages 124–163. Oxford University Press.

Cutler, D. M. and Meara, E. (2000). The Technology of Birth: Is It Worth It? In Garber, A. M., editor, *Frontiers in Health Policy Research, Volume 3*, chapter 2. MIT press, Boston.

Cutler, D. M., Rosen, A. B., and Vijan, S. (2006). The value of medical spending in the United States, 1960-2000. *The New England journal of medicine*, 355(9):920–7.

De-Regil, L. M., Fernández-Gaxiola, A. C., Dowswell, T., and Peña Rosas, J. P. (2010). Effects and safety of periconceptional folate supplementation for preventing birth defects. *The Cochrane database of systematic reviews*, (10):CD007950.

Dehejia, R. and Lleras-Muney, A. (2004). Booms, Busts, and Babies' Health. *The Quarterly Journal of Economics*, 119(3):1091–1130.

Department of Health (2003). Toolkit for High Quality Neonatal Care. Technical report, Department of Health, London.

Department of Health (2011). *Resource Allocation: Weighted Capitation Formula*. Department of Health, London, 7th editio edition.

Department of Health (2012). Health and Social Care Act 2012: fact sheets.

Department of Health (2013a). 2012-13 Programme Budgeting Guidance. Technical report, Department of Health.

Department of Health (2013b). NHS Allocations 2012/12. Technical report, Department of Health, London.

Dolan, P. and Tsuchiya, A. (2006). The elicitation of distributional judgements in the context of economic evaluation. In Jones, A. M., editor, *The Elgar Companion to Health Economics*, pages 382–391. Edward Elgar Publishing, London, 1st editio edition.

Doyle, K. J. and Bradshaw, W. T. (2012). Sixty golden minutes. *Neonatal network : NN*, 31(5):289–94.

Doyle, L. W. (2001). Outcome at 5 Years of Age of Children 23 to 27 Weeks' Gestation: Refining the Prognosis. *PEDIATRICS*, 108(1):134–141.

Ehrenkranz, R. A., Walsh, M. C., Vohr, B. R., Jobe, A. H., Wright, L. L., Fanaroff, A. A., Wrage, L. A., and Poole, K. (2005). Validation of the National Institutes of Health consensus definition of bronchopulmonary dysplasia. *Pediatrics*, 116(6):1353–60.

Fellman, V., Hellström-Westas, L., Norman, M., Westgren, M., Källén, K., Lagercrantz, H., Marsál, K., Serenius, F., and Wennergren, M. (2009). One-year survival of extremely preterm infants after active perinatal care in Sweden. *JAMA : the journal of the American Medical Association*, 301(21):2225–33.

Field, D. and Draper, E. S. (1999). Survival and place of delivery following preterm birth: 1994-96. *Archives of disease in childhood. Fetal and neonatal edition*, 80(2):F111–4.

Filho, F. L., da Silva, A. A. M., Lopes, J. M. A., Lamy, Z. C., Simões, V. M. F., and dos Santos, A. M. (2011). Staff workload and adverse events during mechanical ventilation in neonatal intensive care units. *Jornal de Pediatria*.

Fordham, R., Field, D. J., Hodges, S., Normand, C., Mason, E., Burton, P., Yates, J., and Male, S. (1992). Cost of neonatal care across a regional health authority. *Journal of public health medicine*, 14(2):127–30.

Frederick, S. and Lowenstein, G. (2003). Hedonic Adaptation. In Kahneman, D., Diener, E., and Schwarz, N., editors, *Well-Being: The Foundations of Hedonic Psychology*, chapter 16, pages 302–329. The Russell Sage Foundation, New York, 1st editio edition.

Freedman, S. (2012). Capacity and Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission.

Freedman, S. M. . (2010). *Empirical essays on the economics of neonatal care*. Doctor of philosophy, University of Maryland.

Fuchs, V. R. (1978). The Supply of Surgeons and the Demand for Operations. *The Journal of Human Resources*, 13:35–56.

Fuchs, V. R. (2004). More variation in use of care, more flat-of-the-curve medicine. *Health affairs (Project Hope)*, Suppl Vari:VAR104–7.

Gale, C., Hay, a., Philipp, C., Khan, R., Santhakumaran, S., and Ratnavel, N. (2012a). In-utero transfer is too difficult: results from a prospective study. *Early human development*, 88(3):147–50.

Gale, C., Santhakumaran, S., Nagarajan, S., Statnikov, Y., and Modi, N. (2012b). Impact of managed clinical networks on NHS specialist neonatal services in England: population based study. *Bmj*, 344(apr03 1):e2105–e2105.

Gaynor, M., Seider, H., and Vogt, W. B. (2005). The Volume-Outcome Effect, Scale Economies, and Learning-by-Doing. *The American Economic Review*, 95(2):243–247 CR – Copyright © 2005 American Econom.

Godfrey, L. (1988). *Misspecifications Tests in Econometrics*. Cambridge University Press, Cambridge, 1st edition.

Goldenberg, R. L., Culhane, J. F., Iams, J. D., and Romero, R. (2008). Preterm Birth 1 Epidemiology and causes of preterm birth. pages 75–84.

Goodman, D. C., Fisher, E. S., Little, G. A., Stukel, T. A., Chang, C.-h., and Schoendorf, K. S. (2002). The relation between the availability of neonatal intensive care and neonatal mortality. *The New England journal of medicine*, 346(20):1538–44.

Grandi, C., González, A., and Meritano, J. (2010). [Patient volume, medical and nursing staffing and its relationship with risk-adjusted outcomes of VLBW infants in 15 Neocosur neonatal network NICUs]. *Archivos argentinos de pediatría*, 108(6):499–510.

Gravelle, H. and Smith, D. (2001). Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health economics*, 10(7):587–99.

Gronau, R. (1977). Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited. *Journal of Political Economy*, 85(6):1099–1123.

Gruber, J. (1997). The Consumption Smoothing Benefits of Unemployment Insurance. *The American Economic Review*, 87(1):192–205.

Hall, M., Smith, S., Jackson, J., Perks, E., and Walton, P. (1992). Neonatal nurse practitioners–a view from perfidious Albion? *Archives of disease in childhood*, 67(4 Spec No):458–462.

Hall, R. and Jones, C. (2007). The Value of Life and the Rise in Health Spending. *Quarterly Journal of Economics*, 122(1):39–72.

Halm, E. A., Lee, C., and Chassin, M. R. (2002a). Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of internal medicine*, 137(6):511–20.

Halm, E. a., Lee, C., and Chassin, M. R. (2002b). Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of internal medicine*, 137(6):511–20.

Hamilton, B. H. (2001). Estimating treatment effects in randomized clinical trials with non-compliance: the impact of maternal smoking on birthweight. *Health economics*, 10(5):399–410.

Hamilton, K. E. S. C., Redshaw, M. E., and Tarnow-Mordi, W. (2007). Nurse staffing in relation to risk-adjusted mortality in neonatal care. *Archives of disease in childhood. Fetal and neonatal edition*, 92(2):F99–F103.

Hansen, A. K., Wisborg, K., Uldbjerg, N., and Henriksen, T. B. (2008). Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. *BMJ (Clinical research ed.)*, 336(7635):85–7.

Harrison, E. M., O'Neill, S., Meurs, T. S., Wong, P. L., Duxbury, M., Paterson-Brown, S., Wigmore, S. J., and Garden, O. J. (2012). Hospital volume and patient outcomes after cholecystectomy in Scotland: retrospective, national population based study. *BMJ*, 344(may23 1):e3330–e3330.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153.

Heckman, J. J. and Walker, J. R. (1990). The Third Birth In Sweden. *Journal of Population Economics*, 3(4):235–75.

Heller, G., Richardson, D. K., Schnell, R., Misselwitz, B., Künzel, W., and Schmidt, S. (2002). Are we regionalized enough? Early-neonatal deaths in low-risk births by the size of delivery units in Hesse, Germany 1990-1999. *International journal of epidemiology*, 31(5):1061–8.

Hinchliffe, S. R., Seaton, S. E., Lambert, P. C., Draper, E. S., Field, D. J., and Manktelow, B. N. (2013). Modelling time to death or discharge in neonatal care: an application of competing risks. *Paediatric and perinatal epidemiology*, 27(4):426–33.

Hollingsworth, B. and Parkin, D. (2001). The efficiency of the delivery of neonatal care in the UK. *Journal of public health medicine*, 23(1):47–50.

Horbar, J. D., Badger, G. J., Lewit, E. M., Rogowski, J., and Shiono, P. H. (1997). Hospital and Patient Characteristics Associated With Variation in 28-Day Mortality Rates for Very Low Birth Weight Infants. *Pediatrics*, 99(2):149–156.

Horrace, W. C. and Oaxaca, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90(3):321–327.

Hotz, J. V., Klerman, J. A., and Willis, R. J. (1997). Chapter 7 The economics of fertility in developed countries. In Rosenzwig, M. R. and Stark, O., editors, *Handbook of Population and Family Economics*, pages 275–347. Elsevier B.V., 1st edition.

Huesch, M. D. (2009). Learning by doing, scale effects, or neither? Cardiac surgeons after residency. *Health services research*, 44(6):1960–82.

Iams, J. D., Romero, R., Culhane, J. F., and Goldenberg, R. L. (2008). Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. *Lancet*, 371(9607):164–75.

Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica: Journal of the Econometric Society*, 62(2):467–475.

Information Centre (2012). NHS Maternity Statistics, England: 2011-12.

Ioannides, Y. M. and Loury, L. D. (2004). Job Information Networks, Neighborhood Effects, and Inequality.

Johansson, S., Montgomery, S. M., Ekbom, A., Olausson, P. O., Granath, F., Norman, M., and Cnattingius, S. (2004). Preterm delivery, level of care, and infant death in sweden: a population-based study. *Pediatrics*, 113(5):1230–5.

Jones, L. E., Schoonbroodt, A., and Tertilt, M. (2010). Fertility Theories: Can They Explain the Negative Fertility-Income Relationship? In Shoven, J., editor, *Demography and the Economy*, pages 43–100. University of Chicago Press, Chicago.

Jones, L. E. and Tertilt, M. (2008). An Economic History of Fertility in the United States: 1826 - 1960. In Rupert, P., editor, *Frontiers of Family Economics*, chapter 5, pages 165–230. Emerald Press.

Juárez, S. P. and Merlo, J. (2013). Revisiting the effect of maternal smoking during pregnancy on offspring birthweight: a quasi-experimental sibling analysis in Sweden. *PloS one*, 8(4):e61734.

Kelley-Quon, L. I., Tseng, C.-H., Scott, A., Jen, H. C., Calkins, K. L., and Shew, S. B. (2012). Does hospital transfer predict mortality in very low birth weight infants requiring surgery for necrotizing enterocolitis? *Surgery*, 152(3):337–343.

Kim, S. and Mohtadi, H. (1992). Labor Specialization and Endogenous Growth. *American Economic Review*, 82(2):404–408.

Lake, E. T., Staiger, D., Horbar, J., Cheung, R., Kenny, M. J., Patrick, T., and Rogowski, J. A. (2012). Association between hospital recognition for nursing excellence and outcomes of very low-birth-weight infants. *JAMA : the journal of the American Medical Association*, 307(16):1709–16.

Lasswell, S. M., Barfield, W. D., Rochat, R. W., and Blackmon, L. (2010). Perinatal regionalization for very low-birth-weight and very preterm infants: a meta-analysis. *JAMA : the journal of the American Medical Association*, 304(9):992–1000.

Latini, G., De Felice, C., Giannuzzi, R., and Del Vecchio, A. (2013). Survival rate and prevalence of bronchopulmonary dysplasia in extremely low birth weight infants. *Early human development*, 89 Suppl 1:S69–73.

Le Ray, C., Zeitlin, J., Jarreau, P. H., Bréart, G., and Goffinet, F. (2009). The influence of level of care on admission to neonatal care for babies of low-risk nullipara. *European journal of obstetrics, gynecology, and reproductive biology*, 144(1):21–6.

Leleu, H., Moises, J., and Valdmanis, V. (2012). Optimal productive size of hospital's intensive care units. *International Journal of Production Economics*, 136(2):297–305.

Levitt, S. D., List, J. a., and Syverson, C. (2013). Toward an Understanding of Learning by Doing Evidence from an Automobile Assembly Plant. *Journal of Political Economy*, 121(4):643–681.

Lin, P. W. and Stoll, B. J. (2006). Necrotising enterocolitis. *Lancet*, 368:1271–83.

Lindo, J. M. (2010). Are Children Really Inferior Goods? Evidence from Displacement-driven Income Shocks. *Journal of Human Resources*, 45:301–327.

Lindo, J. M. (2011). Parental job loss and infant health. *Journal of health economics*, 30(5):869–79.

Lippert-Rasmussen, K. (2001). Egalitarianism, Option Luck, and Responsibility.

Lipscomb, J. (1989). Time preference for health in cost-effectiveness analysis. *Medical care*, 27(3 Suppl):S233–S253.

Lorch, S. a., Baiocchi, M., Ahlberg, C. E., and Small, D. S. (2012). The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics*, 130(2):270–8.

Lovenheim, M. F. and Mumford, K. J. (2013). Do Family Wealth Shocks Affect Fertility Choices? Evidence from the Housing Market. *Review of Economics and Statistics*, 95(2):464–475.

Lucas, R. E. (1988). On the mechanics of economic development.

Luce, B. R., Mauskopf, J., Sloan, F. a., Ostermann, J., and Paramore, L. C. (2006). The return on investment in health care: from 1980 to 2000. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 9(3):146–56.

Luft, H. S., Bunker, J. P., and Enthoven, A. C. (1979). Should operations be regionalized? The empirical relation between surgical volume and mortality. *The New England journal of medicine*, 301(25):1364–9.

Luft, H. S., Hunt, S. S., and Maerki, S. C. (1987). The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health services research*, 22(2):157–82.

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., and Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 European countries. *The New England journal of medicine*, 358(23):2468–81.

Madden, C. W. (1999). Excess capacity: markets regulation, and values. *Health services research*, 33(6):1651–1668.

Mangham, L. J., Petrou, S., Doyle, L. W., Draper, E. S., and Marlow, N. (2009). The cost of preterm birth throughout childhood in England and Wales. *Pediatrics*, 123(2):e312–27.

Manktelow, B. N., Seaton, S. E., Field, D. J., and Draper, E. S. (2013). Population-based estimates of in-unit survival for very preterm infants. *Pediatrics*, 131(2):e425–32.

Manning, A. (2009). You can't always get what you want: The impact of the UK Jobseeker's Allowance. *Labour Economics*, 16(3):239–250.

March of Dimes (2006). Global Report on Birth Defects. Technical report.

Marmot, M. (2004). 3 Social Causes of Social Inequalities in Health. In Anand, S., Peter, F., and Sen, A., editors, *Public Health, Ethics, and Equity*, pages 37–63. Oxford University Press, New York, 1st editio edition.

Martin, S., Rice, N., and Smith, P. C. (2008). Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of health economics*, 27(4):826–42.

Medlock, S., Ravelli, A. C. J., Tamminga, P., Mol, B. W. M., and Abu-Hanna, A. (2011). Prediction of mortality in very premature infants: a systematic review of prediction models. *PloS one*, 6(9):e23441.

Menard, M. K., Liu, Q., Holgren, E. A., and Sappenfield, W. M. (1998). Neonatal mortality for very low birth weight deliveries in South Carolina by level of hospital perinatal service. *American journal of obstetrics and gynecology*, 179(2):374–81.

Menclova, A. K. (2012). The Effects of Unemployment on Prenatal Care Use and Infant Health. *Journal of Family and Economic Issues*, (2007).

Mistry, H., Dowie, R., Franklin, R. C. G., and Jani, B. R. (2009). Costs of neonatal care for low-birthweight babies in English hospitals. *Acta paediatrica (Oslo, Norway : 1992)*, 98(7):1123–9.

Monitor (2013). A guide to the Market Forces Factor. Technical report, NHS England, London.

Monitor (2014). Approved Costing Guidance. Technical report, Monitor, London.

Morris, S., Sutton, M., and Gravelle, H. (2005). Inequity and inequality in the use of health care in England: an empirical investigation. *Social science & medicine (1982)*, 60(6):1251–66.

Moster, D., Lie, R. T., and Markestad, T. (1999). Relation between size of delivery unit and neonatal death in low risk deliveries: population based study. *Archives of disease in childhood. Fetal and neonatal edition*, 80(3):F221–5.

Murray, C. J. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization*, 72(3):429–445.

National Audit Office (2007). Caring for Vulnerable Babies: The Reorganisation of neonatal services in England. Technical report, National Audit Office, London.

National Institute for Health and Care Excellence (2013). Guide to the methods of technology appraisal 2013. Technical report, National Institute for Health and Care Excellence, London.

Neto, M. T. (2006). Perinatal care in Portugal: effects of 15 years of a regionalized system. *Acta paediatrica (Oslo, Norway : 1992)*, 95(11):1349–52.

NHS England (2013). Programe Budgeting Finance Manual 2012/13. Technical report, NHS England, London.

Nijman, T. and Verbeek, M. (1992). Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics*, 7(3):243–257.

Noble, M., Mclennan, D., Wilkinson, K., Whitworth, A., Exley, S., Barnes, H., and Dibben, C. (2007). The English indices of deprivation 2007.

Nord, E., Enge, A. U., and Gundersen, V. (2010). QALYs: is the value of treatment proportional to the size of the health gain? *Health economics*, 19(5):596–607.

Nordhaus, W. (2002). The Health of Nations: The Contribution of Improved Health to Living Standards.

Norman, J. E., Morris, C., and Chalmers, J. (2009). The effect of changing patterns of obstetric care in Scotland (1980-2004) on rates of preterm birth and its neonatal consequences: perinatal database study. *PLoS medicine*, 6(9):e1000153.

Northern Neonatal Network (1993). Measuring neonatal nursing workload. Northern Neonatal Network. *Archives of disease in childhood*, 68(5 Spec No):539–43.

Office for National Statistics (2010). Standard Occupational Classification 2010 (SOC2010).

Office for National Statistics (2011). Super Output Areas.

Office for National Statistics (2013). Historic and Projected Mortality Data from the Period and Cohort Life Tables, 2012-based, UK, 1981-2062. Technical report, Office for National Statistics.

Office for National Statistics (2014). Child Mortality Statistics: Childhood, Infant and Perinatal, 2012. Technical report, Office for National Statistics, London.

Office of National Statistics (2011). Birth cohort tables, England and Wales, 2008.

Office of National Statistics (2012). Birth Summary Tables, England and Wales, 2011 (final).

O'Neill, C. and Largey, a. (1997). Issues in cost function specification for neonatal care: the Fordham case. *Journal of public health medicine*, 19(1):50–4.

O'Neill, C., Malek, M., Mugford, M., Normand, C., Tarnow-Mordi, W. O., Hey, E., and Halliday, H. L. (2000). A cost analysis of neonatal care in the UK: results from a multicentre study. ECSURF Study Group. *Journal of public health medicine*, 22(1):108–15.

Papke, L. E. and Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1-2):121–133.

Parfit, D. (1997). Equality and Priority. *Ratio*, 10:202–21.

Parmanum, J., Field, D., Rennie, J., and Steer, P. (2000). National census of availability of neonatal intensive care. *BMJ : British Medical Journal*, 321(7263):727–729.

Payne, R. A. and Able, G. A. (2012). UK indices of multiple deprivation – a way to make comparisons across constituent countries easier. *Health Statistics Quarterly Office For National Statistics*, 53:22–37.

Phibbs, C. S. (2012). Managed clinical networks in neonatal care. *BMJ*, 344(apr03 1):e2423–e2423.

Phibbs, C. S., Baker, L. C., Caughey, A. B., Danielsen, B., Schmitt, S. K., and Phibbs, R. H. (2007). Level and volume of neonatal intensive care and mortality in very-low-birth-weight infants. *The New England journal of medicine*, 356(21):2165–75.

Phibbs, C. S., Bronstein, J. M., Buxton, E., and Phibbs, R. H. (1996). The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *JAMA : the journal of the American Medical Association*, 276(13):1054–9.

Philip, A. G. S. (2005). The evolution of neonatology. *Pediatric research*, 58(4):799–815.

Pillay, T., Nightingale, P., Owen, S., Kirby, D., and Spencer, S. A. (2011). Neonatal nursing efficacy: practical standards of nursing care provision in a newborn network. *Archives of Disease in Childhood*, 96(Supplement 1):A36–A36.

Pollack, M. M. and Koch, M. a. (2003). Association of outcomes with organizational characteristics of neonatal intensive care units. *Critical care medicine*, 31(6):1620–9.

Profit, J., McCormick, M. C., Escobar, G. J., Richardson, D. K., Zheng, Z., Coleman-Phox, K., Roberts, R., and Zupancic, J. a. F. (2007). Neonatal intensive care unit census influences discharge of moderately preterm infants. *Pediatrics*, 119(2):314–9.

Profit, J., Petersen, L. A., McCormick, M. C., Escobar, G. J., Coleman-Phox, K., Zheng, Z., Pietz, K., and Zupancic, J. A. F. (2010). Patient-to-nurse ratios and outcomes of moderately preterm infants. *Pediatrics*, 125(2):320–6.

Profit, J., Zupancic, J. a. F., McCormick, M. C., Richardson, D. K., Escobar, G. J., Tucker, J., Tarnow-Mordi, W., and Parry, G. (2006). Moderately premature infants at Kaiser Permanente Medical Care Program in California are discharged home earlier than their peers in Massachusetts and the United Kingdom. *Archives of disease in childhood. Fetal and neonatal edition*, 91(4):F245–50.

Rautava, L., Lehtonen, L., Peltola, M., Korvenranta, E., Korvenranta, H., Linna, M., Hallman, M., Andersson, S., Gissler, M., Leipälä, J., Tammela, O., and Häkkinen, U. (2007). The effect of birth in secondary- or tertiary-level hospitals in Finland on mortality in very preterm infants: a birth-register study. *Pediatrics*, 119(1):e257–63.

Richter, M., Erhart, M., Vereecken, C. a., Zambon, A., Boyce, W., and Nic Gabhainn, S. (2009). The role of behavioural factors in explaining socio-economic differences in adolescent health: a multilevel study in 33 countries. *Social science & medicine (1982)*, 69(3):396–403.

Roberts, D. and Dalziel, S. (2006). Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *The Cochrane database of systematic reviews*, (3):CD004454.

Roblin, D. W., Richardson, D. K., Thomas, E., Fitzgerald, F., Veintimilla, R., Hulac, P., Bemis, G., and Leon, L. (2000). Variation in the use of alternative levels of hospital care for newborns in a managed care organization. *Health services research*, 34(7):1535–53.

Rogowski, J. a., Horbar, J. D., Staiger, D. O., Kenny, M., Carpenter, J., and Geppert, J. (2004). Indirect vs direct hospital quality indicators for very low-birth-weight infants. *JAMA : the journal of the American Medical Association*, 291(2):202–9.

Royal College of Pediatrics and Child Health (2013). UK-WHO 0-4 years growth chart resources.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Ruud, P. A. (1983). Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models. *Econometrica*, 51(1):225–228.

Saigal, S. and Doyle, L. W. (2008a). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet*, 371(9608):261–9.

Saigal, S. and Doyle, L. W. (2008b). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet*, 371(9608):261–9.

Samuelson, J. L., Buehler, J. W., Norris, D., and Sadek, R. (2002). Maternal characteristics associated with place of delivery and neonatal mortality rates among very-low-birthweight infants, Georgia. *Paediatric and perinatal epidemiology*, 16(4):305–13.

Sanderson, M., Sappenfield, W. M., Jespersen, K. M., Liu, Q., and Baker, S. L. (2000). Association between level of delivery hospital and neonatal outcomes among South Carolina Medicaid recipients. *American journal of obstetrics and gynecology*, 183(6):1504–11.

Schelling, T. (1987). *The New Palgrave: A Dictionary of Economics*. The Macmillian Press Limited, London.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380.

Sen, A. (1985). *Commodities and Capabilities*. Oxford India paperbacks. Oxford University Press.

Sfekas, A. (2009). Learning, forgetting, and hospital quality: an empirical analysis of cardiac procedures in Maryland and Arizona. *Health economics*, 18(6):697–711.

Shah, K. K. (2009). Severity of illness and priority setting in healthcare: a review of the literature. *Health policy (Amsterdam, Netherlands)*, 93(2-3):77–84.

Sherenian, M., Profit, J., Schmidt, B., Suh, S., Xiao, R., Zupancic, J. a. F., and DeMauro, S. B. (2013). Nurse-to-patient ratios and neonatal outcomes: a brief systematic review. *Neonatology*, 104(3):179–83.

Shields, M. A. (2004). Addressing nurse shortages: what can policy makers learn from the econometric evidence on nurse labour supply?*. *The Economic Journal*, 114(499):F464–F498.

Shim, J. W., Kim, M. J., Kim, E.-K., Park, H. K., Song, E. S., Lee, S. M., Lee, J. H., Jin, H.-S., Kim, E. S., and Chang, Y. S. (2013). The impact of neonatal care resources on regional variation in neonatal mortality among very low birthweight infants in Korea. *Paediatric and perinatal epidemiology*, 27(2):216–25.

Smith, L. K., Draper, E. S., Manktelow, B. N., Dorling, J. S., and Field, D. J. (2007). Socioeconomic inequalities in very preterm birth rates. *Archives of disease in childhood. Fetal and neonatal edition*, 92(1):F11–4.

Smith, L. K., Draper, E. S., Manktelow, B. N., and Field, D. J. (2009). Socioeconomic inequalities in survival and provision of neonatal care: population based study of very preterm infants. *Bmj*, 339(dec01 1):b4702–b4702.

Smith, S. L. and Hall, M. a. (2011). Advanced neonatal nurse practitioners in the workforce: a review of the evidence to date. *Archives of disease in childhood. Fetal and neonatal edition*, 96(2):F151–5.

Solans, M., Pane, S., Estrada, M. D., Serra-Sutton, V., Berra, S., Herdman, M., Alonso, J., and Rajmil, L. (2008). Health-related quality of life measurement in children and adolescents: A systematic review of generic and disease-specific instruments.

Stenson, B. J., Tarnow-Mordi, W. O., Darlow, B. A., Simes, J., Juszczak, E., Askie, L., Battin, M., Bowler, U., Broadbent, R., Cairns, P., Davis, P. G., Deshpande, S., Donoghoe, M., Doyle, L., Fleck, B. W., Ghadge, A., Hague, W., Halliday, H. L., Hewson, M., King, A., Kirby, A., Marlow, N., Meyer, M., Morley, C., Simmer, K., Tin, W., Wardle, S. P., and Brocklehurst, P. (2013). Oxygen saturation and outcomes in preterm infants. *The New England journal of medicine*, 368(22):2094–104.

Stevenson, D. K. (2000). Sex differences in outcomes of very low birthweight infants: the newborn male disadvantage. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 83(3):182F–185.

Stone, M. (1979). Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):276–278.

Straney, L. D., Lim, S. S., and Murray, C. J. L. (2012). Disentangling the effects of risk factors and clinical care on subnational variation in early neonatal mortality in the United States. *PloS one*, 7(11):e49399.

Stringhini, S., Sabia, S., Shipley, M., Brunner, E., Nabi, H., Kivimaki, M., and Singh-Manoux, A. (2010). Association of socioeconomic position with health behaviors and mortality. *JAMA : the journal of the American Medical Association*, 303(12):1159–66.

Stukel, T. A., Lucas, F. L., and Wennberg, D. E. (2005). Long-term Outcomes of Regional Variations in Intensity of Invasive vs Medical With Acute Myocardial Infarction. 293(11):1329–1337.

Synnes, A. R., Macnab, Y. C., Qiu, Z., Ohlsson, A., Gustafson, P., Dean, C. B., and Lee, S. K. (2006). Neonatal intensive care unit characteristics affect the incidence of severe intraventricular hemorrhage. *Medical care*, 44(8):754–9.

Talbot-Smith, A. and Pollock, A. (2006). *The New NHS: A Guide: A Guide to Its Funding, Organisation and Accountability*. Routledge, London, 1st editio edition.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, 27(3):531–43.

Thomas, S., Fayter, D., Misso, K., Ogilvie, D., Petticrew, M., Sowden, A., Whitehead, M., and Worthy, G. (2008). Population tobacco control interventions and their effects on social inequalities in smoking: systematic review. *Tobacco control*, 17(4):230–7.

Towers, C. V., Bonebrake, R., Padilla, G., and Rumney, P. (2000). The effect of transport on the rate of severe intraventricular hemorrhage in very low birth weight infants. *Obstetrics and gynecology*, 95(2):291–5.

Tucker, J. (2002). Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. *Lancet*, 359(9301):99–107.

Usher, R. (1971). Clinical Implications of Perinatal Mortality Statistics. *Clinical Obstetrics & Gynecology*, 14(3):885–925.

van den Berg, G. J., Lindeboom, M., and Portrait, F. (2006). Economic Conditions Early in Life and Individual Mortality. *American Economic Review*, 96(1):290–302.

Van Reempts, P., Gortner, L., Milligan, D., Cuttini, M., Petrou, S., Agostino, R., Field, D., den Ouden, L., Bø rch, K., Mazela, J., Carrapato, M., and Zeitlin, J. (2007). Characteristics of neonatal units that care for very preterm infants in Europe: results from the MOSAIC study. *Pediatrics*, 120(4):e815–25.

Varni, J. W., Limbers, C. A., Neighbors, K., Schulz, K., Lieu, J. E. C., Heffer, R. W., Tuzinkiewicz, K., Mangione-Smith, R., Zimmerman, J. J., and Alonso, E. M. (2011). The PedsQL Infant Scales: Feasibility, internal consistency reliability, and validity in healthy and ill infants. *Quality of Life Research*, 20(1):45–55.

Vermont-Oxford Network (2014). Vermont-Oxford Network.

Wall, S. N., Handler, A. S., and Park, C. G. (2004). Hospital factors and nontransfer of small babies: a marker of deregionalized perinatal care? *Journal of perinatology : official journal of the California Perinatal Association*, 24(6):351–9.

Watson, S. I., Arulampalam, W., Petrou, S., Marlow, N., Morgan, a. S., Draper, E. S., Santhakumaran, S., and Modi, N. (2014). The effects of designation and volume of neonatal care on mortality and morbidity outcomes of very preterm infants in England: retrospective population-based cohort study. *BMJ Open*, 4(7):e004856–e004856.

Wilcox, A. J. (2001). On the importance–and the unimportance– of birthweight. *International Journal of Epidemiology*, 30(6):1233–1241.

Williams, S., Whelan, a., Weindling, a. M., and Cooke, R. W. (1993). Nursing staff requirements for neonatal intensive care. *Archives of disease in childhood*, 68(5 Spec No):534–8.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1):115–132.

Wooldridge, J. M. (2003). Chapter 17. In *Econometric Analysis of Cross Section and Panel Data*, pages 623–625. MIT Press, 4th printi edition.

Woolridge, J. M. (2010). Correlated random effects models with unbalanced panels.

Yang, X. and Borland, J. (1991). A Microeconomic Mechanism for Economic Growth.

Zeitlin, J. and Ancel, P.-y. (2011). Interpreting data on the health outcomes of extremely preterm babies. *Archives of disease in childhood. Fetal and neonatal edition*, 96(5):F314–6.

Ziliak, J. P. (1997). Predetermined : Empirical of Estimators Comparison. *Journal of Business & Economic Statistics*, 15(4):419–431.

Appendix A

Addition Information for Chapter 3: Literature Review

A.1 Table of Results

Table A.1 Results from the literature review

Author	Year	Country	Study population	Study methods	Outcome measures	Results	Notes
Abdel-Latif et al.	2006	Australia	10 units, n=8654, <32 weeks ga, 1992-2002	Multivariate logistic regression	7 day mortality, categorized by time of admission.	36.7% of infants were admitted after hours on a weekdays and 28.4% on weekends and public holidays. Admission time has little impact on outcomes, the only recorded statistically significant difference being between CS delivery rates	Evening and weekend admissions were analysed as staff levels are lower during these times
Almond, Doyle, Jr., Kowalski, & Williams	2010	US	NCHS birth cohort 1983-1991, 1995-2002 giving 30-200,000 births depending on analysis	Regression discontinuity analysis of births just above and just below 1500g.	Mortality and costs associated with band just above 1500g against just below to determine returns to increased spending on medical care	Mortality is lower by about 1% just below 1500g compared to just above, spending is about $4,000 higher just below the threshold leading to estimate of around $550,000/life saved in 2002 dollars. A newborn life is worth about $2.7m (Cutler and Meara 2000) so very cost-effective	

Baker & Phibbs	2002	California, US		Hazard models for adoption of NICU (mid- and high-level), logistic regression for mortality, multinomial logistic regression	HMO market share and diffusion of NICUs. 28 day mortality.	There was a reduction in mortality odds of 38-45% for infants born in high level units as opposed to mid level units. If there are more mid-level units locally then fewer babies end up in high level NICUs. As HMOs gain greater market share they become less likely to set up mid level NICUs as they are less likely to adopt new technologies, this increases survival as more infants are transferred to high level hospitals	The proximity to certain level units influences outcomes in that area.
Bartels et al.	2006	Germany	36 NICUs (107 delivery units), n=4379, 24-30 weeks	Univariate logistic regression based on general estimating equations	28 day mortality	The odds ratio of mortality in a small NICU within 28 days was 1.23 (95% CI: 1.01-1.49). The same relationship was found when including many other covariates. The relationship was true for NICUs regardless of the size of the delivery unit.	The findings here reflect those of many other studies.
Bell et al.	2010	Iowa, US	11137 babies, 501-1250g, 2001-5	Poisson regression	7 and 28 day mortality, morbidities (BPD, IVH, PVL, ROP)	no effect of the timing of birth on mortality rate and no impact on the risks of short-term morbidities except that the risk of retinopathy of prematurity (stage 2) was higher after the introduction of duty-hour restrictions and the risk of retinopathy of prematurity requiring operative treatment was lower for infants born during the late night than during the day.	

Bode, O'shea, Metzguer, & Stiles	2001	South Carolina, US	69,452 babies, 500-2000g, 1968-1994	multivariate logistic regression models, with separate models for birth location and stratification by birth weight	28 day mortality	More babies have been born in level III units over time, the number of level III units has also increased. The risk of dying has is decreased at level III hospitals after 1974.	The regressions did not control for many common diseases and abnormalities to give a true estimate or relative risk.
Callaghan et al.	2003	Australia	692 VLBW in one unit, 1996-1999	Multiple logistic regression, ORs	Survival to discharge	There were 80 deaths among the 692 babies analysed for the study period. The odds of mortality adjusted for initial risk and infant dependency scores (unit workload), were improved by 82% when an infant/staff ratio of greater than 1.71 occurred, suggesting improved survival with the highest infant/staff ratio. The low and medium staffing levels corresponded with similar odds ratios for mortality.	Study has converse findings to most similar studies on infant/staff ratios.

Cimiotti et al.	2006	New York, US	2,675 NICU admissions in two units, 2001-3	Cox proportional hazards model	Time to first episode of healthcare-associated bloodstream infection.	A total of 224 infants had an infection that met the study definition of healthcare-associated bloodstream infection. In a multivariate analysis, after controlling for infants' intrinsic and extrinsic risk factors, a greater number of hours of care provided by registered nurses in NICU 2 was associated with a decreased risk of bloodstream infection in these infants (hazard ratio, 0.21; 95% confidence interval, 0.06-0.79).	
Judith H Chung et al.	2010	California, US	38,313 VLBW babies in 1997-2002	Logistic regression	28 day and 1 year mortality	Higher volume units were associated with a lower death rate, although the trend was not linear. Low birth weight along with a number of maternal characteristics increased risk of death.	This study is very similar to Chung et al. 2011 but doesn't utilise as advanced modelling techniques. The authors write that deregionalisation in California has not led to more infants being born in higher level units, and infants are generally being transferred to mid level units from regional centres.

J H Chung et al.	2011	California, US	167 (of 177) units in 1997-2002	Hierarchical generalised linear modelling	28 day and 1 year mortality	Crude death rates were lower in high volume hospitals. The highest odds of death were in units with an annual volume of 10 or fewer very low birth weight deliveries. Specifically, these units had 80% higher odds (odd ratio 1.79; 95% confidence interval 1.32–2.42) of very low birth weight death when compared with the highest volume units (>100 very low birth weight deliveries, annually). The presence of training programs did not seem to affect death rates.	The authors note that previous studies have not used multilevel modelling whereas they do. Most previous studies haven't gone beyond looking at delivery volume or level of care. Doesn't control for confounders, the only difference is in volume, not level.
Cifuentes et al.	2002	California, US	16,732 <2000g in 1992-1993	Logistic regression	28 day mortality (incl. deaths within 1 year if infant remained hospitalised)	NICU size is associated with greater odds of mortality, an effect amplified by birth weight.	This analysis is fairly crude compared to other subsequent studies on the topic
Field & Draper	1999	UK	16 (Trent health region): <32 weeks ga 1994-1996	Various statistical methods (including Stu- dent's t test, Wilcoxon two tailed sample test and chi sq. , as appropriate) were used	Mortality	Larger units had significantly more smaller, more immature, lower 5 min Apgar, multiple births, breech births and respiratory distress infants. Mortality rates between smaller and larger units were not significantly different.	It is unclear whether this analysis correctly controlled for unit level variables when considering outcomes. Moreover, they had fairly limited data.

Filho et al.	2011	Brazil	543 admissions to NICUs	Generalised estimating equations	Adverse events	The larger the number of newborns classified by care demand (NCCD) per nurse and nursing technician, the more likely the occurrence of intermediate adverse events linked to mechanical ventilation. A number of NCCD > 22 per nurse (relative risk [RR] = 2.86) and > 4.8 per auxiliary nurse (RR = 3.41) was associated with a higher prevalence of intermediate adverse events.	
Fordham et al.	1992	UK	Births in Trent RHA between 1987-1988	Cost data were obtained using three-stage postal questionnaire and telephone follow up	Costs		This is the first paper to try and determine neonatal costs. There have been subsequent, updated and improved versions, see O'Neill et al 2000

Goodman et al.	2002	US	246 Regions, 1995 US birth cohort >500g, n=3,892,208	Logistic regression	27 day mortality	In general there was no consistent assoication between the supply of NICU beds or neonatologists and neonatal mortality. However there was a small difference in risk between regions with a very low supply of neonatologists and a low supply. High levels of neonatologists were found in areas where the mothers had the highest rate of multiple births and primiparous mothers but the lowest rates of motherless than 12 years education	There are significant differences in the US in supply of neonatologists and NICU beds. The resources in the US are maldistributed. However study lacks control for many individual and unit level variables.
Grandi et al.	2010	South America	2,019 VLBW in 15 NICUs, 2005-7	Multilevel logistic regression	Mortality,severe IVH, BPD, ROP, sepsis.	A low medical hours provision was significantly associated with increased mortality (OR 1.30 [95% CI: 1.04-1.76], p= 0.020); on the other hand low nurse provision was significantly associated with increased risk of mortality, adjusted by mother age and initial risk (trained NIC 1.52 [1.16 -1.99], nurses-to-infant ratio 1.81 [1.40-2.33]). Although public centers showed higher risk of morbidity and mortality compared with private centers, differences were statistically not significant.	

Hamilton et al.	2007	UK	54: 2,636 babies <1500g, <31 ga	(Multivariate) logistic regression	Mortality	The average shift was understaffed (calculated as a function of number of admissions in each shift). Risk adjusted mortality was not different for infants born in low volume units. Mortality decreased with increased specialist nursing staff ratio (although it appears not significantly)	This paper is unclear on how is controlled for important covariates and the analysis seems fairly crude. Morbidity is not examined here.
Hollingsworth and Parkin	2001	UK	49 neonatal units	Data envelopment analysis	Unit efficiency and costs	Economies of scale varied between units, with increasing returns in the 36 inefficient units, and mainly constant returns in the 13 efficient units. This suggests that the presence of technical inefficiency was as important as scale inefficiencies. Total cost savings, if all units were operating efficiently, are estimated at £10.4 million, equivalent to 10 extra units producing 57,000 additional days of care.	This was a re-analysis of the work by O'Neill et al. (2000).
Horbar, Badger, Lewit, Rogowski, & Shiono	1997	Vermont Oxford network	7672 babies 501-1500g , 1991-2, 62 NICUs	Logistic regression with cluster robust standard errors	28 day mortality and standardised mortality ratios	There was no association between annual patient volume and either mortality rate (r=.17) or SNMR (r=.22). Observed mortality rates (17% vs 13%) and SNMR (1.04 vs .87) were both higher at the 24 hospitals with pediatric residency training programs. However in models adjusting for both volume and training program, neither estimate was significant	

Lake et al.	2012	US	72235 VLBW babies between 2007-8 in 558 Vermont Oxford network units	Multilevel logistic regression	7 and 28 day mortality and IVH and nosocomial infection	The 7-day mortality was 7.0% in RNE hospitals and 7.4% in non- RNE hospitals (adjusted odds ratio [OR], 0.87; 95% CI, 0.76-0.99; P=.04). The 28- day mortality was 10.0% in RNE hospitals and 10.5% in non-RNE hospitals (adjusted OR, 0.90;95%CI, 0.80-1.01; P=.08). Hospital stay mortality was 12.4% in RNE hos- pitals and 13.1% in non-RNE hospitals (adjusted OR, 0.90;95%CI, 0.81-1.01; P=.06). Severe intraventricular hemorrhage was 7.2% in RNE hospitals and 7.8% in non-RNE hospitals (adjusted OR, 0.88; 95% CI, 0.77-1.00; P=.045). Infection was 16.7% in RNE hospitals and 18.3% in non-RNE hospitals (adjusted OR, 0.86; 95% CI, 0.75- 0.99; P=.04).	RNE = recognition for nursing excellence
Leleu, Moises, & Valdmanis	2012	Florida, US	51 NICUs	Non-parametric estimation of returns to scale using linear program	Number of care days	The average number of beds of a NICU under constant returns to scale was 38, the average nurse hours per bed was 10.4 – this was less than hospitals under increasing (15.6) and decreasing (19.9) returns to scale.	This is the only study to examine returns to scale and to model a production technology for neonatal units.

Lorch et al.	2012	Pennsylvania, California, and Missouri, US	1,328132 hospital based deliveries, 1995-2005	IV approach.	Mortality, BPD, infection	Infants who were delivered at a high-level NICU had significantly fewer in-hospital deaths in Pennsylvania (7.8 fewer deaths/1000 deliveries, 95% confidence interval [CI] 4.1–11.5), California (2.7 fewer deaths/1000 deliveries, 95% CI 0.9–4.5), and Missouri (12.6 fewer deaths/1000 deliveries, 95% CI 2.6–22.6). Deliveries at high-level NICUs had similar rates of most complications, with the exception of lower bronchopulmonary dysplasia rates at Missouri high-level NICUs (9.5 fewer cases/1000 deliveries, 95% CI 0.7–18.4) and higher infection rates at high-level NICUs in Pennsylvania and California. The association between delivery hospital, in-hospital mortality, and complications differed across the 3 states.	
O'Neill & Largey	1997	UK	Births in Trent RHA between 1987-1988	Linear regression	Costs	The average costs per day were sensitive to model specification.	See O'Neill et al. 2000 for an update

O'Neill et al.	2000	UK	57	Linear regression (A double-log function relating variations in total costs to total days, case-mix and an interaction term provided the best fit to the data)	Costs	Four models were considered. There are economies of scale present, for example an increase from 2000 to 3000 days of care, for units with proportions of intensive care days of 0.01, 0.02, 0.05 and 0.1 produced, respectively, a 27, 23, 19 and 16 per cent fall in the average daily cost of care.	Scale effects are more noticeable in units providing relatively less intensive care. This is probably in part due to less opportunity for labour substitution
Phibbs et al.	2007	California, US	48,237, 501-1500g in 1991-2000	Logistic regression	28 day mortality (incl. deaths within 1 year for infants who remain hospitalised)	Deliveries at hospitals with lower-level and lower-volume NICUs were associated with an increased risk of death, adjusted for the risk factors considered. By considering those deliveries that could have ocurred in higher level hospitals the authors estimate that 21% of the deaths of very-low-birthweight infants in the year 2000 were potentially preventable	The authors write that it is difficult to separate the effects of level and volume in assessing unit outcomes. Here they note that it was not possible to assess whether there was a causal relationship between NICU characteristics and outcomes.

Phibbs et al.	1996	California, US	594,104 Births in non-federal hospitals with maternity services	Multivariate logistic regression	Compared with hospitals without an NICU, infants born in a hospital with a level III NICU with an average NICU census of at least 15 patients per day had significantly lower risk-adjusted neonatal mortality (odds ratio, 0.62; 95% confidence interval, 0.47-0.82; P=.002). Risk-adjusted neonatal mortality for infants born in smaller level III NICUs, and in level II+ and level II NICUs, regardless of size, was not significantly different from hospitals without an NICU, and was significantly higher than hospitals with large level III NICUs.

Pollack & Koch	2003	Washington DC, US	552 babies in 8 units, 500-1499g, 199-1997	Questionnaires for unit organisational characteristics, univariate and multivariate logistic and linear models	28 day mortality, morbidities (BPD, PIVH/PVL, ROP)	A relationship was found between leadership, coordination, conflict resolution and morbidity (specifically for PIVH/PVL). When staff rated these attributes of their hospital higher they found that the prevalence of PIVH/PVH was lower	The authors note that few studies up to that point had addressed specific structure and processes of ICUs. Of interest, one previous paper had found that daily rounds of by an ICU physician lead to a three fold lower risk for many complications and morbidity and mortality.
Profit et al.	2010	California and Massachusetts, US	850 moderately preterm in 10 NICUs	Multilevel logistic regression	Daily weight gain, mortality, gestational age at discharge, IVH, BPD, sepsis	In bivariate analyses, an increase in PNR was associated with a slightly higher daily weight gain (5 g/day), greater gestational age at discharge, any intraventricular hemorrhage, and severe retinopathy of prematurity. After controlling for case mix, NICU size, and site of care, an additional patient per nurse was associated with a decrease in daily weight gain by 24%. Other variables were no longer independently associated with PNR.	

Roblin et al.	2000	California, US	4 (Kaiser Permanente hospitals)	Likelihood is estimated by logistic regression, duration of stay is estimated by Cox proportional hazards	Length of stay	The likelihood of admission into NICU and the duration of both NICU care and hospital stay are proportional to the degree ofillness and complexity of diagnosis. Admission does not seem to be affected by method of reimbursement. The difference is also not accounted for by supply of beds.
Rogowski et al.	2004	US	94,110 births between 1995-2000, 501-1500g, Vermont Oxford Network	Logistic regressions	Mortality prior to discharge	Higher volume units had a reduction in mortality (OR 0.989 CI: 0.983-0.994) up to a volume >50 infants when there is no longer a difference. Lower level units had higher mortality and did more urban ones. However volume does not explain much of the variability in unit mortality (9%).
Shim et al.	2013	Korea	2584 VLBW in 91 neonatal units, 2009	Adjusted logistic regression	Mortality	Infants who were delivered at a low-level NICU had an increased odd of mortality (level 1 OR: 1.7, 95% CI: 1.2-2.5; level 2 OR: 1.8, 95% CI: 1.3-2.5) compared to level three units. A higher odds of mortality was observed for infants delivered at a NICU with 20-49 annual VLBW admissions compared to infants delivered at a NICU with 50+ VLBW admissions (OR: 1.7, 95% CI: 1.3-2.3).

Straney et al	2012	US	44015582 live births between 1996-2006	Adjusted logistic regression	Early neonatal mortality	Accounting for preterm volume (defined as ,34 weeks), the number of neonatologist and NICU beds, 25.2% and 58.7% of the HSA-level variance in outcomes was explained. Odds of early neonatal mortality were lower for infants born in areas with a NICU, with greater numbers of obstetritions per 1,000 births, greater numbers of neonatologists per 1,000 births, and with higher volume NICUs.	
Synnes et al.	2006	Canada	17 units, <33 weeks ga, <4 days old at admission, 1996-7, n=3772	Univariate and bivariate Bayesian hierarchical logistic regression	IVH	Larger NICU patient volume (odds ratio 0.7, 95% credible interval 0.6–0.9) and higher neonatologist-housestaff ratio (odds ratio 0.2, 95% credible interval 0.0–0.7) were associated with a reduced risk of IVH. NICU characteristics account for 31% of the variation in the incidence of severe IVH.	Use of Bayesian hierarchical regression to model outcomes, but final model was random intercept model. The only model to examine morbidity and staffing. NICU characteristics account for as much of the variation of severe IVH as do population baseline characteristics.

Tucker	2002	UK	12		Hospital mortality and nosocomial bacteracaemia	Detailed data of staff costs per bed day and per bed. Relationship between staffing and patient length of stay. There was not significant effect of extended nursing roles and quality of care or outcomes.	The report is the results of the UK PICU staffing survey. This report contains data relating to a wide range of staffing characteristics including costs, wellbeing etc.
Wall et al.	2004	Illinois, US	2904 babies in 10 units, 1989-1996, <1250g	Multiple logistic regression, ORs	Transfer status	The transfer rate was higher among lower level hospitals. Birth rate was associated with risk of transfer but not linearly. Hospitals with higher than average volume of VLBW and with increased revenues from Medicaid and HMO were more likely to transfer and so were hospitals attached to networks.	This study shows that there are many factors associated with nontransfer, not just hospital level. This has an implication for clinical and economic outcomes.

Appendix B

Addition Information for Chapter 4: Neonatal Unit Volume and Designation

B.1 Regression Results

This section contains the full regression results for the main analyses presented in chapter 4. Table B.1 shows the results for the standard logistic regression for very preterm ($\leq 32^{+6}$) infants admitted to high volume neonatal unit at the place of birth. Table B.2 shows the equivalent standard logistic regression results for admission to tertiary level neonatal unit at the place of birth. Tables B.3 and B.4 show the results from the equivalent instrumental variables logistic regressions respectively.

Table B.1 Parameter estimates from standard logistic regression for admission to high volume neonatal unit at the place of birth for very preterm infants

	(1) Neonatal mortality	(2) Any in hospital mortality	(3) BPD	(4) Treatment for ROP	(5) Surgery for NEC	(6) PMA at discharge >40 weeks
Hi volume	-0.321*	-0.190	0.108	-0.0534	0.0501	0.125
	(0.136)	(0.124)	(0.0700)	(0.170)	(0.162)	(0.0969)
Gestational age	-2.328***	-2.069***	6.779***	1.820*	1.083	1.508***
	(0.319)	(0.303)	(0.295)	(0.711)	(0.594)	(0.293)
Gestational age sq.	0.0328***	0.0279***	-0.133***	-0.0458***	-0.0264*	-0.0357***
	(0.00570)	(0.00544)	(0.00520)	(0.0137)	(0.0109)	(0.00525)
2008	-0.181	-0.106	0.0763	0.0549	0.0269	0.135
	(0.0926)	(0.0776)	(0.0574)	(0.177)	(0.167)	(0.0723)
2009	-0.253**	-0.208*	0.0342	0.349*	-0.180	0.167*
	(0.0820)	(0.0815)	(0.0511)	(0.136)	(0.145)	(0.0772)
Z-score	-0.129*	-0.239***	-0.421***	-0.449***	-0.436***	-0.711***
	(0.0538)	(0.0476)	(0.0271)	(0.0718)	(0.0709)	(0.0331)
Antenatal ster.	-0.292***	-0.227**	-0.0257	0.0969	-0.0459	0.0260
	(0.0864)	(0.0770)	(0.0419)	(0.123)	(0.128)	(0.0611)
Bottom 10% dep.	0.160	0.132	0.0151	0.234	-0.313	0.124
	(0.109)	(0.0941)	(0.0760)	(0.151)	(0.177)	(0.0638)
Male sex.	0.137	0.251***	0.387***	0.116	0.288*	0.434***
	(0.0747)	(0.0662)	(0.0445)	(0.133)	(0.125)	(0.0557)
Constant	36.46***	33.16***	-85.10***	-19.57*	-13.66	-16.50***
	(4.409)	(4.185)	(4.204)	(9.168)	(8.057)	(4.057)
N	20,554	20,554	20,554	20,554	20,554	20,554

Values are parameter estimates (cluster robust standard errors in parentheses)

* p<0.05; ** p<0.01; *** p<0.001

BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birthweight z score (Z-score), use of antenatal steroids (antenatal ster.), male sex, infant year of birth and deprivation decile.

High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Table B.2 Parameter estimates from standard logistic regression for admission to level three neonatal unit at the place of birth for very preterm infants

	(1) Neonatal mortality	(2) Any in hospital mortality	(3) BPD	(4) Treatment for ROP	(5) Surgery for NEC	(6) PMA at discharge >40 weeks
Level 3	-0.260	-0.0942	0.204**	0.234	0.0493	0.157
	(0.133)	(0.119)	(0.0684)	(0.166)	(0.162)	(0.0954)
Gestational age	-2.324***	-2.069***	6.788***	1.737*	1.082	1.505***
	(0.315)	(0.301)	(0.295)	(0.700)	(0.593)	(0.292)
Gestational age sq.	0.0327***	0.0279***	-0.133***	-0.0441**	-0.0264*	-0.0356***
	(0.00564)	(0.00541)	(0.00518)	(0.0135)	(0.0108)	(0.00524)
2009	-0.194*	-0.111	0.0833	0.0660	0.0292	0.142*
	(0.0953)	(0.0793)	(0.0574)	(0.175)	(0.167)	(0.0715)
2010	-0.259**	-0.210*	0.0378	0.356**	-0.179	0.169*
	(0.0826)	(0.0817)	(0.0511)	(0.136)	(0.145)	(0.0771)
Z-score	-0.129*	-0.236***	-0.419***	-0.439***	-0.436***	-0.709***
	(0.0543)	(0.0480)	(0.0272)	(0.0711)	(0.0711)	(0.0328)
Antenatal ster.	-0.283**	-0.225**	-0.0324	0.0771	-0.0482	0.0195
	(0.0871)	(0.0772)	(0.0415)	(0.125)	(0.128)	(0.0605)
Bottom 10% dep.	0.183	0.139	-0.00902	0.211	-0.318	0.107
	(0.109)	(0.0943)	(0.0721)	(0.151)	(0.178)	(0.0613)
Male sex	0.139	0.251***	0.385***	0.112	0.288*	0.432***
	(0.0753)	(0.0666)	(0.0443)	(0.134)	(0.125)	(0.0558)
Constant	36.36***	33.08***	-85.33***	-18.77*	-13.64	-16.50***
	(4.350)	(4.153)	(4.195)	(9.038)	(8.064)	(4.044)
N	20,554	20,554	20,554	20,554	20,554	20,554

Values are parameter estimates (cluster robust standard errors in parentheses)
* p<0.05; ** p<0.01; *** p<0.001
BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birthweight z score (Z-score), use of antenatal steroids (antenatal ster.), male sex, infant year of birth and deprivation decile.

Table B.3 Parameter estimates from instrumental variable logistic regression for admission to high volume neonatal unit at the place of birth for very preterm infants

	(1) Neonatal mortality	(2) Any in hospital mortality	(3) BPD	(4) Treatment for ROP	(5) Surgery for NEC	(6) PMA at discharge >40 weeks
Hi volume	-0.361*	-0.390***	0.0471	0.0207	0.228	-0.0840
	(0.141)	(0.115)	(0.105)	(0.270)	(0.255)	(0.123)
Gestational age	-2.332***	-2.089***	6.771***	1.825*	1.102	1.485***
	(0.319)	(0.304)	(0.296)	(0.710)	(0.595)	(0.295)
Gestational age sq.	0.0329***	0.0281***	-0.133***	-0.0459***	-0.0266*	-0.0354***
	(0.00570)	(0.00545)	(0.00520)	(0.0137)	(0.0109)	(0.00528)
2009	-0.179	-0.0975	0.0785	0.0515	0.0191	0.144*
	(0.0936)	(0.0781)	(0.0574)	(0.178)	(0.167)	(0.0709)
2010	-0.252**	-0.206*	0.0350	0.348*	-0.182	0.169*
	(0.0820)	(0.0811)	(0.0512)	(0.137)	(0.145)	(0.0774)
Z-score	-0.130*	-0.243***	-0.423***	-0.448***	-0.433***	-0.715***
	(0.0548)	(0.0482)	(0.0273)	(0.0730)	(0.0708)	(0.0332)
Antenatal ster.	-0.292***	-0.227**	-0.0258	0.0970	-0.0457	0.0257
	(0.0863)	(0.0764)	(0.0419)	(0.123)	(0.128)	(0.0616)
Bottom 10% dep.	0.162	0.143	0.0182	0.230	-0.324	0.136*
	(0.109)	(0.0941)	(0.0757)	(0.151)	(0.176)	(0.0654)
Male sex	0.137	0.251***	0.387***	0.117	0.289*	0.435***
	(0.0749)	(0.0664)	(0.0444)	(0.133)	(0.125)	(0.0556)
Residual	0.0584	0.290	0.0950	-0.104	-0.263	0.309*
	(0.246)	(0.216)	(0.0937)	(0.256)	(0.279)	(0.133)
Constant	36.55***	33.62***	-84.94***	-19.72*	-14.08	-15.99***
	(4.417)	(4.198)	(4.218)	(9.177)	(8.086)	(4.089)
N	20,554	20,554	20,554	20,554	20,554	20,554

Values are parameter estimates (cluster robust standard errors in parentheses)

* p<0.05; ** p<0.01; *** p<0.001

BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birthweight z score (Z-score), use of antenatal steroids (antenatal ster.), male sex, infant year of birth, deprivation decile, and first stage generalised residual (Resid.)

High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Table B.4 Parameter estimates from instrumental variable logistic regression for admission to level three neonatal unit at the place of birth for very preterm infants

	(1) Neonatal mortality	(2) Any in hospital mortality	(3) BPD	(4) Treatment for ROP	(5) Surgery for NEC	(6) PMA at discharge >40 weeks
Level 3	-0.137	-0.164	0.173	0.647*	0.157	-0.0561
	(0.142)	(0.116)	(0.114)	(0.254)	(0.247)	(0.130)
Gestational age	-2.309***	-2.079***	6.784***	1.789*	1.097	1.475***
	(0.315)	(0.301)	(0.297)	(0.698)	(0.599)	(0.295)
Gestational age sq.	0.0326***	0.0281***	-0.133***	-0.0448***	-0.0266*	-0.0352***
	(0.00564)	(0.00541)	(0.00520)	(0.0135)	(0.0109)	(0.00528)
2009	-0.195*	-0.111	0.0830	0.0583	0.0286	0.142*
	(0.0955)	(0.0795)	(0.0573)	(0.177)	(0.168)	(0.0711)
2010	-0.261**	-0.208*	0.0378	0.346*	-0.180	0.171*
	(0.0828)	(0.0818)	(0.0512)	(0.138)	(0.145)	(0.0774)
Z-score	-0.126*	-0.237***	-0.420***	-0.433***	-0.434***	-0.713***
	(0.0552)	(0.0487)	(0.0275)	(0.0718)	(0.0711)	(0.0331)
Antenatal ster.	-0.286**	-0.223**	-0.0317	0.0665	-0.0510	0.0248
	(0.0873)	(0.0774)	(0.0413)	(0.125)	(0.128)	(0.0608)
Bottom 10% dep.	0.167	0.149	-0.00433	0.156	-0.334	0.139*
	(0.106)	(0.0926)	(0.0728)	(0.147)	(0.176)	(0.0642)
Male sex	0.139	0.251***	0.385***	0.117	0.288*	0.432***
	(0.0751)	(0.0666)	(0.0443)	(0.135)	(0.125)	(0.0558)
Resid.	-0.182	0.103	0.0489	-0.586*	-0.161	0.323*
	(0.231)	(0.200)	(0.117)	(0.273)	(0.286)	(0.131)
Constant	36.04***	33.28***	-85.24***	-19.89*	-13.96	-15.87***
	(4.351)	(4.159)	(4.240)	(9.020)	(8.191)	(4.094)
N	20,554	20,554	20,554	20,554	20,554	20,554

Values are parameter estimates (cluster robust standard errors in parentheses)

* p<0.05; ** p<0.01; *** p<0.001

BPD=Bronchopulmonary Dysplasia, PMA at discharge=postmenstrual age at discharge, equal to gestational age at birth plus the length of stay in weeks. Models are adjusted for gestational age, gestational age squared, birthweight z score (Z-score), use of antenatal steroids (antenatal ster.), male sex, infant year of birth, deprivation decile, and first stage generalised residual (Resid.)

High volume was defined as being in the top quartile of units by number of care days provided to infants born at $\leq 32^{+6}$ weeks gestation.

Appendix C

Addition Information for Chapter 5: Neonatal Healthcare Expenditure

C.1 A Simple Example of the Issue With Aggregate Expenditure Data

A simple example is provided here to explain the issues with the use of aggregated data to estimate the effects of healthcare expenditure of patient clinical outcomes explicated in chapter 5. Consider three individuals, A, B, and C each of whom may contract an illness, the clinical symptom of which is a cough. Medicine is available that can treat the illness, the success of which is measured by a reduction in the coughing rate. We are interested in identifying the cost-effectiveness of the medicine by determining the average cough reduction per pound sterling spent on medicine. We only observe the total expenditure for all three individuals and the total number of coughs in a given period. The cost-effectiveness of the medicine is therefore estimated by comparing differing levels of medical expenditure on the total number of coughs.

Table C.1 shows three alternative scenarios with varying spending on medicine and the number of coughs. A comparison between scenario 1 and scenario 2 would reveal the cost-effectiveness of the medicine since we are comparing the same two individuals. In this case it would be 1.5 coughs reduced per pound sterling. Comparing scenario

Table C.1 Expenditure on medicine and the number of coughs in three different scenarios

	Scenario 1		Scenario 2		Scenario 3	
	Medicine (£)	Coughs	Medicine (£)	Coughs	Medicine (£)	Coughs
A	10	20	5	30	10	20
B	20	25	15	30	20	25
C	0	0	0	0	10	22.5
Total	30	45	20	60	40	67.5

1 and scenario 3, however, would not reveal the effects of medical spending, since the cause of the change in aggregate expenditure is due to a change in the identity of the individuals receiving medicine. Comparing scenarios 1 and 3 would give us an estimated cost-effectiveness of -1 coughs reduced per person. Thus, at an aggregate level, we require data on expenditure and outcomes, for the same individuals. Typically, aggregate expenditure data will incorporate both types of changes, those arising from different medical inputs and those arising due to a change in the patient population.

To isolate the effects of medical expenditure using aggregate data we can utilise some exogenous variable that affects the level of medical expenditure. However, to be exogenous in the coughing example, this variable would have to influence the spending on medicine but be independent of the risk of developing the illness.

In the Claxton et al. (2013) study the instrumental variables for healthcare expenditure that are utilised are socio-economic variables. However, it was argued in chapter 5, that these instruments are not independent of the factors that determine the identity of the patients that are admitted to healthcare services. Moreover, if these factors that determine the risk of developing an illness, are independent of the clinical outcomes post admission, then these instruments will appear to be valid in tests of instrument validity, such as a test of overidentifying restrictions.

C.2 Regression Results

This section presents the full regression results from chapter 5. Table C.2 shows the regression results treating expenditure as exogenous, and table C.3 shows the equivalent results treating expenditure as endogenous.

Table C.2 Regression results treating expenditure as exogenous

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
ex	−0.000225	−0.0160***	−0.0668***
	(−0.20)	(−7.80)	(−8.63)
Gestational age	−0.139***	−0.672***	−1.159
	(−13.10)	(−17.13)	(−1.68)
Gestational age sq.	0.00185***	0.0109***	0.0203
	(12.34)	(16.50)	(1.45)
Antenatal ster.	−0.0246***	−0.0242***	−0.0609*
	(−7.66)	(−4.27)	(−2.14)
Sex	−0.00231***	−0.0229***	−0.0992***
	(−4.04)	(−3.63)	(−4.50)
Birth weight z	−0.00408***	−0.0218***	−0.0777***
	(−9.43)	(−6.59)	(−6.27)
Dep quin. 2	−0.0384	−0.101	0.0555
	(−1.60)	(−1.25)	(1.81)
Dep quin. 3	−0.0388	−0.105	0.0228
	(−1.61)	(−1.28)	(0.72)
Dep quin. 4	−0.0384	−0.106	0.0385
	(−1.60)	(−1.29)	(1.24)
Dep quin. 5	−0.0382	−0.0928	0.108**
	(−1.61)	(−1.27)	(2.79)
2009/10	−0.00551*	−0.0543***	−0.268***
	(−2.14)	(−7.70)	(−9.58)
2010/11	0	0.0137	0.0295
	(.)	(1.62)	(0.73)
2011/12	−0.00136		
	(−1.64)		
2012/13	−0.0137	−0.0922	0.662
	(−0.68)	(−0.65)	(0.85)
Multiple	−0.00679***	−0.00893	0.0361
	(−4.31)	(−1.10)	(1.30)
MFF	0.155	1.009	−7.016
	(0.71)	(0.65)	(−0.79)
Constant	2.507***	9.742***	24.87*
	(8.20)	(5.84)	(2.10)
N	101,559	12,777	1,729

[1] * p<0.05; ** p<0.01; *** p<0.001; Cluster robust standard errors in parentheses

[2] The dependent variable is in-hospital mortality.

[3] The control variables are gestational age, gestational age squared, birth weight z-score (Birth weight z), indicators for whether a full or partial course of antenatal steroids was administered and male sex (Antenatal ster.), year fixed effects, deprivation quintile dummies (Dep. quin.), and place of birth fixed effects.

[4] IC=infants who received at least one day of intensive care. $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.

Table C.3 Regression results treating expenditure as endogenous

	(1) Whole sample	(2) $\leq 32^{+6}$	(3) $\leq 26^{+6}$
ex	−0.00219**	−0.0227***	−0.0720***
	(0.000744)	(0.00198)	(0.00604)
Gestational age	−0.144***	−0.667***	−1.092
	(0.00592)	(0.0366)	(0.630)
Gestational age sq.	0.00192***	0.0109***	0.0189
	(0.0000814)	(0.000618)	(0.0128)
Antenatal ster.	−0.0261***	−0.0243***	−0.0557*
	(0.00193)	(0.00537)	(0.0271)
Gender	−0.00225***	−0.0238***	−0.0938***
	(0.000535)	(0.00611)	(0.0176)
Birth weight z	−0.00411***	−0.0223***	−0.0733***
	(0.000401)	(0.00304)	(0.0110)
Dep. quin. 2	−0.0332	−0.103	0.0655*
	(0.0203)	(0.0598)	(0.0265)
Dep. quin. 3	−0.0335	−0.105	0.0227
	(0.0205)	(0.0599)	(0.0208)
Dep. quin. 4	−0.0329	−0.107	0.0511
	(0.0206)	(0.0603)	(0.0264)
Dep. quin. 5	−0.0327	−0.0939	0.106**
	(0.0201)	(0.0550)	(0.0380)
2009/10	0.00655	0.0372	−0.941
	(0.0195)	(0.131)	(0.766)
2010/11	0.0147	0.113	−0.638
	(0.0193)	(0.129)	(0.759)
2011/12	0.0131	0.101	−0.663
	(0.0194)	(0.132)	(0.760)
Multiple	−0.00657***	−0.00933	0.0457
	(0.00139)	(0.00527)	(0.0246)
MFF	0.153	1.061	−7.122
	(0.210)	(1.449)	(8.743)
N	101,559	12,776	1,719
J statistic	0.761	0.897	0.694
J-stat p-value	0.102	0.118	0.461

[1] * p<0.05; ** p<0.01; *** p<0.001. Cluster robust standard errors in parentheses

[2] The dependent variable is in-hospital mortality.

[3] The control variables are gestational age, gestational age squared, birth weight z-score (Birth weight z), indicators for whether a full or partial course of antenatal steroids was administered and male sex (Antenatal ster.), year fixed effects, Index of Multiple Deprivation quintile dummies (Dep. quin.), and place of birth fixed effects.

[4] IC=infants who received at least one day of intensive care. $\leq 32^{+6}$=infants born at $\leq 32^{+6}$ weeks gestation. $\leq 26^{+6}$=infants born at $\leq 26^{+6}$ weeks gestation.

Appendix D

Addition Information for Chapter 7: One to One Nursing

D.1 Regression Results

This section presents the full regression results from the main analyses presented in chapter 6.

Table D.1 Estimated effect of one to one nursing rate on mortality rate

One to one nursing variable defined as the proportion of intensive care days on which one to one nursing was provided

	(1) OLS	(2) 2SGMM	(3) 2SLS	(4) 2SGMM
1:1 Nursing	0.0566*	-0.0562*	-0.0420	-0.0436*
	(0.0223)	(0.0282)	(0.0317)	(0.0201)
2008	-0.00884*	-0.0113**	-0.0133**	-0.00831+
	(0.00430)	(0.00436)	(0.00470)	(0.00485)
2009		-0.00880**	-0.00990*	-0.00590+
		(0.00335)	(0.00412)	(0.00325)
2010	0.0175**	0.0113**	0.00951*	0.0101*
	(0.00513)	(0.00420)	(0.00447)	(0.00433)
2011	0.0153**	0.00757+	0.00619	0.0127***
	(0.00528)	(0.00433)	(0.00453)	(0.00347)
2012	0.00897+			
	(0.00508)			
Mean gestational age	-0.00625***	-0.00610***	-0.00627***	-0.00552***
	(0.000925)	(0.000813)	(0.000856)	(0.000843)
Mean z-score	-0.00147	-0.00484	-0.00484	0.00124
	(0.00562)	(0.00527)	(0.00547)	(0.00375)
Proportion antenatal ster.	-0.0173+	-0.0261*	-0.0273*	-0.0167
	(0.0103)	(0.0120)	(0.0122)	(0.0114)
Proportion male	0.0223*	0.0212*	0.0222*	0.0176+
	(0.00938)	(0.00916)	(0.00939)	(0.0107)
Proportion multiple	-0.00367	-0.00850	-0.00549	-0.0140
	(0.0156)	(0.0127)	(0.0131)	(0.0149)
Proportion bottom dep. quin.	-0.00444+	0.00268	0.00294	0.00195
	(0.00236)	(0.00279)	(0.00300)	(0.00254)
need	0.000631	0.000672	0.000370	0.000811
	(0.00115)	(0.00102)	(0.00107)	(0.000855)
Constant	0.247***			
	(0.0319)			
AIC	-6174.1	-6308.9	-6319.6	-5269.4
RMSE	0.0574	0.0560	0.0559	0.0472
Hansen J-statistic		1.440	1.440	0.449
Hansen J-statisticp		0.696	0.696	0.930
N	2,149	2,149	2,149	1,610

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Column (1) is estimated using an OLS estimator (OLS) and treats nursing as exogenous. Columns (2) and (4) are estimated by two stage GMM (2SGMM), and column (3) by two stage least sqaures (2SLS). Columns (1)-(3) use data from level three units while column (4) uses data from high volume units. The dependent variable in each case is the mortality rate (measured between zero and one). Estimates are interpreted as the percentage point changing in the mortality rate resulting from a one percentage point increase in the proportion of intensive care days with one to one nursing. Regressions control for the mean values of gestational age, birth weight z-score, antenatal steroid receipt, gender, and deprivation quintile. RMSE = root mean squared error.

Appendix E

Addition Information for Chapter 7: Economic Conditions and Infant Health

E.1 Further Details on the Econometric Specification

E.1.1 Model 3: Correction for Sample Selection Bias

This analysis follows the method proposed by Wooldridge (1995). The notation below is as described in chapter 7. The selection equation is

$$s_{it} = 1(\delta_0 + z'_{it}\delta_1 + \delta_2 * \text{Unemp.}_{it} + x'_{it}\delta_3 + \xi_i + v_{it} > 0) \tag{E.1}$$

where z_{it} is the previously specified instrumental variable, v_{it} is an $N(0,1)$ random variable, and ξ_i is area unobserved heterogeneity.

Estimation of model (7.2) then proceeds as follows. Letting $x^+_{it} = [z_{it}, \text{Unemp.}_{it}, x_{it}]'$:

1. Estimate equation (E.1) by standard probit separately for each time period, letting $\xi_i = \gamma_0 + \bar{x}^{+'}_{it}\gamma_1 + u_{3it}$, and obtain the inverse Mill's ratio, $\hat{\lambda}_{it}$.

2. Model the unobserved heterogeneity in equation (7.1) as

$$\alpha_i = \theta_0 + \theta_1 \text{Unemp.} + x_{it}' \theta_2 + u_{\alpha,i} \tag{E.2}$$

and let $w_{it} = [x_{i1}^+, ..., x_{iT}^+, x_{it}, 0, ..., 0, \hat{\lambda}_{it}, 0, ..., 0]$ then estimate $y_{it} = w_{it}' \psi + \text{error}_{it}$ by pooled OLS.